University of Alberta

Peak-to-Average Power Ratio Reduction in OFDM Systems

by

Luqing Wang ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

Department of Electrical and Computer Engineering

Edmonton, Alberta
Spring 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-45619-4
Our file Notre référence
ISBN: 978-0-494-45619-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Orthogonal Frequency Division Multiplexing (OFDM) suffers from high Peak-to-Average Power Ratio (PAR) issues. A large PAR requires a linear High Power Amplifier (HPA), which, however, is inefficiently used. Moreover, the combination of an HPA having insufficient linear-range and a large PAR leads to in-band distortion and out-of-band radiation. Various PAR reduction techniques have been proposed.

This thesis focuses mainly on the PAR reduction of OFDM systems by using tone-reservation, sign-selection, and coding techniques. We analyzed the clipping noise by approximating it as a series of parabolic pulses. Our analysis explains peak regrowth and the constant clipping noise power spectrum over the whole OFDM band. We also establish the roughly proportional relationship between the clipping noise at the end of several clipping and filtering iterations, and that generated in the first iteration. The constant of proportionality is estimated via the level-crossing theory. Two algorithms are proposed to reduce the PAR under the tone-reservation constraints. These algorithms scale the filtered first-iteration clipping noise by a constant or adaptively-calculated factor to compensate for peaks above the threshold. The complexity analysis and simulation show that our proposed algorithms achieve a larger PAR reduction and lower complexity than the active-set algorithm.

The PAR can also be reduced by optimizing the signs of data symbols. The optimal signs may be transmitted to the receiver as side information to correctly decode data symbols. In this thesis, we propose an adaptive mapping scheme to eliminate the need for side information at the receiver. We also propose several algorithms (based on using a stochastic search or clipping noise as a guide) to solve the associated discrete optimization problem of the sign-selection technique. The complexity analysis and simulation confirm the complexity advantages of the proposed algorithms compared to the selective mapping and the derandomization algorithms.

Finally, we generalize the Rudin-Shapiro sequence. Constructed from an initial Phase Shift Keying (PSK) sequence, the generalized sequence increases the coding rate at the cost of an increased PAR. By optimizing the signs of the initial sequence, the PAR of the generalized sequence can be further reduced.

To my wife and both our families.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Dr. Chintha Tellambura. Throughout the years, his invaluable advice, continuous encouragement, and constructive criticism guided me to learn and enjoy the art of research. It was my privilege and pleasure to be a member of his team.

I would also like to thank my colleagues and friends, Ms. Yu Fu, Mr. Ling Zhao, Mr. Sa Li, Mr. Lin Fu, and many others. I have always enjoyed my lively discussions with them on a variety of topics. Their friendship has given me so much happiness and helped me to enjoy my life.

My Ph.D. studies were supported by the Department of Electrical and Computer Engineering, the Province of Alberta Graduate Fellowship, and NSERC. Their financial support is greatly appreciated.

I would like express my deepest gratitude to my wife, Wenxin Zhang, and our families for their endless support and love. During every minute of my life, they are always the source of my power and courage.

Table of Contents

1 Introduction 1

1.1 A Brief Overview of OFDM 1

1.2 Structure and Contributions of the Thesis 4

2 OFDM Systems 7

2.1 OFDM Signalling 7

2.2 Modulation and Demodulation Procedure 9

2.3 Cyclic Prefix and FEQ 10

2.4 Peak-to-Average Power Ratio 11

3 PAR-Reduction Techniques 14

3.1 High-Power Amplifiers 14

3.1.1 Soft Limiter 14

3.1.2 Solid State Power Amplifier 15

3.1.3 Traveling-Wave Tube 15

3.1.4 PAR Distribution and BER Performance 15

3.2 PAR-Reduction Techniques 19

3.3 Categorization of the PAR Problem 19

3.3.1 PAR-Reduction Techniques with Distortion 20

3.3.2 Tone-Reservation 23

3.3.3 Active Constellation Extension 27

3.3.4 Probabilistic Techniques 28

3.3.5 Coding Techniques 36

3.3.6 PAR Reduction for MIMO OFDM Systems 38

4 PAR Reduction Using Clipping-based Techniques **39**

4.1 Problem Formulation 39

4.2 Analysis of Clipping and Filtering 41

4.2.1 Time Domain Analysis of Clipping Noise 41

4.2.2 Frequency Domain Analysis of Clipping Noise 45

4.2.3 Clipping Noise Power Spectral Density 47

4.2.4 Filtered Clipping Noise 49

4.2.5 Iterative Clipping and Filtering: Single Clipping Pulse 51

4.2.6 Iterative Clipping and Filtering: Multiple Pulses 54

4.2.7 Effect of Reserved-Tone Position on PAR Reduction 56

4.3 New Tone-Reservation Algorithms 61

4.3.1 Constant Scaling Tone Reservation Algorithm 61

4.3.2 Adaptive-Scaling Tone-Reservation Algorithm 63

4.3.3 Complexity Comparison 65

4.4 Simulation Results 69

4.4.1 Theoretical and Actual Values of β 69

4.4.2 Performance of Proposed Algorithms 73

4.4.3 Tradeoff of PAR Reduction and Execution Time for the Adaptive-Scaling Algorithm 79

4.5 The Constant-Scaling Algorithm: when Inband Distortion Is Allowed 80

4.6 Conclusions 84

4.7 Appendices 85

4.7.A Time Independence of Different Clipping Pulses 85

4.7.B The Conditional pdf and moments of γ_i 87

4.7.C $\Pr[\dot{r}(t_i) > 0 | \dot{r}(t_i) = 0, r(t_i) \geq A] \to 0$ when $A \to \infty$ 90

4.7.D Given $\dot{r} = 0$ and $r \geq A$, γ_i and τ_i Are Uncorrelated 92

4.7.E The Conditional pdf and moments of η_k 93

4.8 The Execution Time of DIT Split-Radix JN-Point FFT with N_f Nonzero Inputs and N Desired Outputs 94

5 PAR Reduction with Discrete Solutions **97**

5.1 Characterization of Discrete Solutions 97

5.2 PAR Reduction By Using Adaptive Mapping 99
5.3 Probabilistic Solutions 102
5.3.1 Random Selection Method 102
5.3.2 Modified PTS Algorithms 103
5.3.3 Recursive Partial Sequence Method 103
5.3.4 Simulation Results 105
5.4 Cross-Entropy Method for PAR Reduction 108
5.4.1 The CE Method 109
5.4.2 The CESS Algorithm for PAR Reduction 112
5.4.3 Modified CE Sign-Selection Algorithm with Threshold 114
5.4.4 Convergence and Complexity Discussion 115
5.4.5 Numerical Results 119
5.5 Clipping Guided Sign-Selection Algorithm for PAR Reduction 128
5.5.1 Problem Formulation 128
5.5.2 Clipping-Noise Guided Sign-Selection Algorithms 129
5.5.3 Complexity Analysis 139
5.5.4 Simulation Results 140
5.6 Conclusions 145

6 Coding Technique **146**
6.1 Rudin-Shapiro Sequence for MPSK Symbols 146
6.2 Further Generalization of Rudin-Shapiro Sequence 147
6.3 Simulation Results 148
6.4 Conclusions 149

7 Conclusions and Future Work **150**
7.1 Conclusions 150
7.2 Future Work 151

Bibliography **155**

List of Tables

4.1 Number of real multiplications/divisions for constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where N=512, $J = 4$. 74

4.2 Average power increase in dB of constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where the reserved tone set is randomly selected. 76

4.3 Given that a clipping pulse occurs in $(0, 4.5T/N)$, the probability that more than k ($k \geq 1$) clipping pulses occur in the same time interval $(0, 4.5T/N)$. 86

4.4 $m_{|\gamma|}$ and $\sigma_{|\gamma|}$ for different A. Simulated results are obtained with $N = 512$, $J = 128$ and QPSK input symbols. 90

5.1 The total numbers of IFFTs for different K 106

5.2 Simulation Parameters for MCESST 124

5.3 Averaged Number of IFFTs required by MCESST 125

5.4 Simulation Parameters for MCESST when 91-Hex is used 127

5.5 Averaged Number of Samples Required by MCESST and SLM when 91-Hex is used 127

5.6 Averaged running time of CGS, SLM, PTS and the derandomization algorithm, where $N = 256$, and 64QAM symbol input is used. 142

List of Figures

1.1 Subchannel division. 2

1.2 Carrier frequency offset and intercarrier interference. 3

2.1 OFDM spectrum of zero-insertion and zero-padding schemes 9

2.2 The block diagram of the OFDM system 10

2.3 PAR distribution for different oversampling factors, $N = 128$. 13

2.4 PAR distribution for different N, $L = 4$. 13

3.1 AM/AM functions of SL, SSPA (for $p = 3$ and $p = 10$), and TWT. . . 16

3.2 AM/PM function of TWT. 16

3.3 Power spectral density of unclipped and time-domain-clipped OFDM signals. 18

3.4 Categorization of the PAR problem 20

3.5 Peak regrowth. 21

3.6 Polygonal approximation of the peak boundary. 26

3.7 Feasible extension region for 16-QAM constellation. 27

3.8 The 91-point hexagonal constellation. 34

3.9 Tone Injection. 34

4.1 Frequency spectrum of $f_i(t)$. 46

4.2 The clipping pulse $f_i(t)$ and its filtered version $\hat{f}_i(t)$. 50

4.3 The probability that more than one clipping pulses occur with a time duration of T_4, given a clipping pulse has already occurred in this time interval, where $\sigma = 1/\sqrt{2}$. 57

4.4 Tone positions and the corresponding impulse response $h(t)$. 60

4.5 Comparison of $\bar{\beta}_{\text{Theo}}$ and $\bar{\beta}_{\text{Simu}}$. 70

4.6 Relative difference d between $\bar{\beta}_{\text{Theo}}$ and $\bar{\beta}_{\text{Simu}}$. 71
4.7 Phase of β_{Simu} and its standard deviation. 72
4.8 Magnitude of β_{Simu} and its standard deviation. 72
4.9 Approximation error D and its standard deviation when using β_{Simu}. 73
4.10 PAR reduction of constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where $R = 4.88\%$, $A = 6.22\,\text{dB}$, and $\mathcal{R}$ is randomly selected. 75
4.11 PAR reduction of constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where $R = 19.92\%$, $A = 4.96\,\text{dB}$, and $\mathcal{R}$ is randomly selected. 76
4.12 BER comparison of adaptive-scaling (AS) and active-set (Act. Set), where $R = 19.92\%$, $A = 4.96\,\text{dB}$, $\mathcal{R}$ is randomly selected, and $C = 5.46\,\text{dB}$. 78
4.13 Out-of-band radiation of adaptive-scaling (AS) and active-set (Act. Set), where $R = 19.92\%$, $A = 4.96\,\text{dB}$, $\mathcal{R}$ is randomly selected, and $C = 5.46\,\text{dB}$. 78
4.14 PAR reduction of adaptive-scaling at the first 16 iterations, where $R = 19.92\%$, $A = 4.96\,\text{dB}$, and $\mathcal{R}$ is randomly selected. 79
4.15 PAR reduction of adaptive-scaling at the first 16 iterations, where $R = 4.88\%$, $A = 6.22\,\text{dB}$, and $\mathcal{R}$ is randomly selected. 80
4.16 PAR CCDF comparison of the constant-scaling algorithm and the iterative clipping and filtering algorithm. 82
4.17 BER performance comparison of the constant-scaling algorithm and the iterative clipping and filtering algorithm. 83
4.18 Out-of-band radiation comparison of the constant-scaling algorithm and the iterative clipping and filtering algorithm for the 3 dB clipping case. 84

5.1 QPSK mapping to 8PSK 100
5.2 PAR for uncoded BPSK symbol sequences and for adaptive-mapping coded symbol sequences, where $N = 16$, $M = 2$. 102

5.3 PAR CCDF for (1) original $\mathcal{Q}_{2M}^{(1)}$ symbol sequences (– –), (2) original $\mathcal{Q}_{2M}$ symbol sequences (——), (3) $\mathcal{Q}_{2M}^{(1)}$ symbol sequences with PTS (– ·), and (4) $\mathcal{Q}_{2M}$ symbol sequences with PTS (· · ·), where $N = 64$, $M = 8$, and each OFDM block is partitioned into four subgroups when using PTS. 104
5.4 PAR comparison, where $N = 256$, $M = 8$. 106
5.5 Basic idea of the CE method. 110
5.6 Relationship between averaged PAR and the averaged number of IFFTs required by CESS for different n, where $N = 128$ and 16QAM symbols are used. 120
5.7 Averaged PAR v.s. the averaged number of IFFTs required by CESS for $n = 128$ and different ρ. 121
5.8 PAR reduction comparison of CESS, SLM and the derandomization method, where $N = 128$ and 16QAM symbols are used. 121
5.9 PAR reduction comparison of CESS and SLM for the same execution time, where 16QAM symbols are used. 123
5.10 PAR reduction comparison of MCESST and SLM for different thresholds, where 16QAM symbols are used. 124
5.11 PAR reduction comparison of CESS, SLM and the derandomization method with $N = 128$ and the 91-Hex constellation. 126
5.12 PAR reduction comparison of CESS and SLM for the same execution time with the 91-Hex constellation. 126
5.13 OFDM system with clipping-noise guided sign-selection. 129
5.14 Clipping-noise guided sign-selection algorithm. 130
5.15 PAR reduction comparison of CGS with $\bar{\beta}$ and adaptive size of $\mathcal{S}$, SLM, PTS, derandomization and tone reservation, where $N = 256$, and 64QAM symbol input is used. 141
5.16 Power spectrum density of SLM, PTS, derandomization and CGS, where $N = 256$, and 64QAM symbol input is used. 143
5.17 PAR reduction comparison of different configurations of CGS, where $N = 256$, and 16QAM symbol input is used. 144

5.18 PAR reduction comparison of CGS with $L = 8$ and different A, where $N = 256$, and 16QAM symbol input is used. 144
5.19 PAR reduction comparison of CGS and SLM for different N, where $K = 16$ for SLM, $L = 8$ and $A = 4\,\text{dB}$ for CGS, and 16QAM symbol input is used. 145

6.1 PAR CCDF of the generalized Rudin-Shapiro sequence for different m. 149

Acronyms

A/D Analog-to-Digital Conversion

ACE Active Constellation Extension

ADSL Asymmetrical Digital Subscriber Line

AWGN Additive White Gaussian Noise

BER bit-error rate

BPSK Binary Phase Shift Keying

CARI cross-antenna rotation and inversion

CCDF Complementary Cumulative Distribution Function

cdf cumulative distribution function

CE Cross-Entropy method

CESS CE Sign-Selection algorithm

CGS clipping-noise guided sign-selection algorithm

CP cyclic prefix

MCESST Modified CE Sign-Selection Algorithm with Threshold

D/A Digital-to-Analog Conversion

DCT Discrete Cosine Transform

DIT decimation-in-time

DFT Discrete Fourier Transform

DMT Discrete Multitone

DWT Discrete Wavelet Transform

FACE Full Adaptive CE method

FEC Forward Error Correction

FEQ Frequency-Domain Equalizer

FFT Fast Fourier Transform

FLOPs float-point operations

HPA High Power Amplifier

IBO Input Back-off

ICI Intercarrier Interference

IDFT Inverse Discrete Fourier Transform

IFFT Inverse Fast Fourier Transform

i.i.d. independent, identically distributed

ISI Inter-Symbol Interference

MAC multiplier-accumulator

MCM Multicarrier Modulation

MIMO Multiple-Input Multiple-Output

ML maximum likelihood

OFDM Orthogonal Frequency Division Multiplexing

P/S Parallel-to-Serial Conversion

PAR Peak-to-Average Power Ratio

pdf probability density function

PTS Partial Transmit Sequences

PSD Power Spectral Density

PSK Phase Shift Keying

QCQP Quadratically Constrained Quadratic Program

QPSK Quadrature Phase Shift Keying

QAM Quadrature amplitude modulation

S/P Serial-to-Parallel Conversion

SFBC Space-Frequency Block Code

SISO Single-Input Single-Output

SL Soft Limiter

SLM Selective Mapping

SNDR Signal-to-Noise-plus-Distortion Ratio

SNR Signal-to-Noise Ratio

SSPA Solid State Power Amplifier

STBC Space-Time Block Code

STD standard deviation

TWT Traveling-Wave Tube

VDSL Very-high-speed Digital Subscriber Line

Notation

- Bold letter: A matrix or a vector.
- A: The clipping threshold.
- N: Number of subcarriers.
- J: The oversampling factor.
- L: Number of iterations.
- R: The tone reservation ratio.
- $\mathcal{R}$: The reserved tone set.
- $\mathcal{Q}_M$: The M-ary signal constellation.
- q_m, $m = 1, 2, ..., M$: A constellation point in $\mathcal{Q}_M$.
- X_k: The modulated data symbol on the k-th subcarrier.
- $\mathbf{X}$: OFDM block where $\mathbf{X} = [X_0, X_1, ..., X_{N-1}]$.
- $x(t)$: The continuous-time domain OFDM signal.
- x_n: The discrete-time domain OFDM signal.
- C_k: The peak reduction signal of the k-th subcarrier.
- $\mathcal{C}$: The signal space of all possible $\mathbf{C}$ vectors.
- f_n, $n = 0, 1, ..., JN - 1$: clipping noise.
- F_k: The frequency spectrum of clipping noise on the k-th subcarrier.

- s_k: The symbol sign of the k-th subcarrier, where $s_k \in \{1, -1\}$.
- $\mathcal{S}$: The index set of data symbols or signal samples.
- T_s: The time duration of a data symbol.
- T: The time duration of an OFDM block.
- β: The scaling factor.
- σ^2: The variance of the real or imaginary part of X_k.
- $E\{\cdot\}$: Expectation.
- $\Re[x]$: The real part of x.
- $\Im[x]$: The imaginary part of x.
- $(\cdot)^*$: Complex conjugate.
- $\langle\cdot\rangle_M$: Modulo-M.
- $\odot$: Element-wise multiplication.

Chapter 1

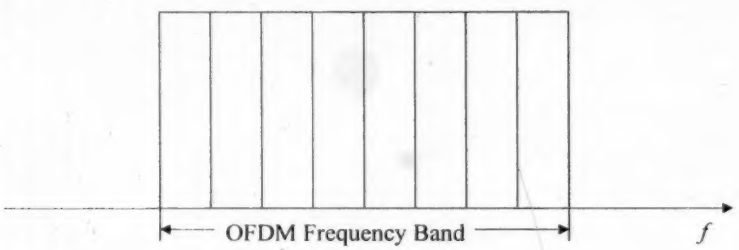
Introduction

1.1 A Brief Overview of OFDM

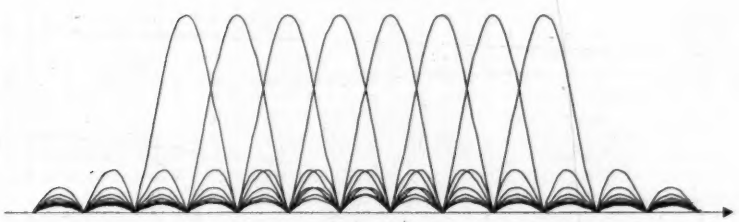
With the development of modern electrical and computer technologies, the demand for fast and reliably transmitting multimedia information over wired or wireless channels is increasing rapidly. High-speed communications must efficiently use a bandlimited channel to obtain a high bit-rate and must combat channel noise, distortion, fading, etc., to maintain a low bit-error rate (BER). Orthogonal Frequency Division Multiplexing (OFDM) is a promising Multicarrier Modulation (MCM) technique for high-speed communications. It has been widely used in a number of communication systems such as IEEE 802.11a/g, IEEE 802.16e, HIPERLAN/2, and Digital Video Broadcasting (DVB). Its baseband version (Discrete Multitone (DMT)) has become the standard modulation technique for the Asymmetrical Digital Subscriber Line (ADSL) and the Very-high-speed Digital Subscriber Line (VDSL). OFDM is also a candidate for IEEE 802.20. A prominent capability of OFDM systems is that it can possibly reach the channel capacity over frequency-selective fading channels.

Unlike the traditional single-carrier modulation techniques, OFDM uses a set of orthogonal subcarriers to divide the communication channel into subchannels. Data bits are transmitted in these subchannels in parallel so as to obtain a high bit-rate [1]. Fig. 1.1(a) shows the ideal "brick wall" subchannel division, where the gain of each subchannel is a constant in its passband and zero elsewhere. Such a subchannel division is, however, unrealistic because it requires infinite-length subchannel filters. Shown in Fig. 1.1(b), OFDM uses finite-length orthogonal sinusoid subcarriers to divide the subchannels. Although the subchannels are overlapped, and the frequency

response of each subchannel has nonzero sidelobes, the gains of the neighboring subchannels are zero at the peak of each subchannel. Thus, the transmitted data symbols can be demodulated at the receiver without Intercarrier Interference (ICI). Usually, a small number of subchannels at the two ends of the OFDM band are not used, in order to prevent interference between neighboring frequency bands.



(a) "Brick wall" subchannel division



(b) OFDM subchannel division

Figure 1.1: Subchannel division.

OFDM can be efficiently implemented by using the Inverse Fast Fourier Transform (IFFT) and the Fast Fourier Transform (FFT). OFDM eliminates the Inter-Symbol Interference (ISI) by using the cyclic prefix. Because the bandwidth of each subchannel is narrow, the subchannel response is approximately constant. Therefore, channel distortion can be combated by using a set of one-tap Frequency-Domain Equalizers (FEQs). The channel equalization in single-carrier systems is more complicated. OFDM is also robust to the narrow-band interference. In wireline communications, the water-filling algorithm can be used to maximize the transmission bit-rate [2]. With this algorithm, subchannels having large Signal-to-Noise Ratios (SNRs) are assigned more bits, and those having small SNRs are assigned less bits or even not used if the SNR is too small. In wireless communications, Forward Error Correction (FEC) codes can be used to combat the narrow-band interference.

The main drawbacks of OFDM are its sensitivity to the ICI (caused by carrier frequency mismatch between the transmitter and receiver, and/or the Doppler shift) and the large Peak-to-Average Power Ratio (PAR) [1]. Fig. 1.2 illustrates the ICI. Because of the carrier frequency offset, the receiver cannot sample the received signal (in the frequency domain) at the frequencies of the peaks of the subchannel responses. The nonzero sidelobes of neighboring subchannels lead to the ICI. In OFDM systems, the largest sidelobe is only 13 dB lower than the peak of the subchannel frequency response. Thus, a small frequency offset may cause intolerable ICI. The ICI problem can be alleviated by using the Discrete Cosine Transform (DCT), the Discrete Wavelet Transform (DWT), or other filterbank-based MCM systems that have small subchannel sidelobes. However, these systems are more complicated than OFDM systems using FFT. Many ICI-cancellation techniques for OFDM also have been proposed in the literature [3–11]. Some PAR-reduction algorithms can also be used to reduce the PAR and ICI simultaneously [6].

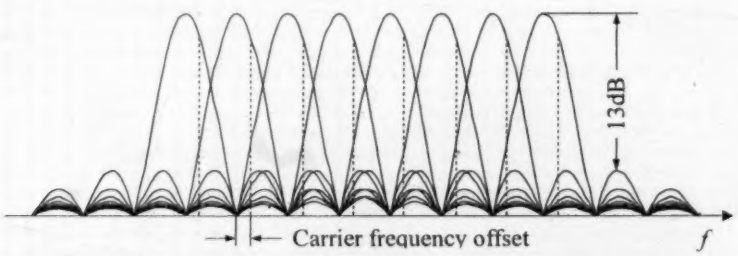


Figure 1.2: Carrier frequency offset and intercarrier interference.

The time-domain baseband equivalent OFDM signal has a complex Gaussian distribution. Thus, compared to its average power, the peak power of the OFDM signal can be very high, but occurs rarely [1]. A large PAR requires a linear High Power Amplifier (HPA), which, however, is inefficiently used. Moreover, the combination of an HPA having insufficient linear-range and a large PAR leads to in-band distortion and out-of-band radiation, where the former increases the BER and the latter interferes with communications in neighboring frequency bands [1]. Various techniques have been proposed to reduce the PAR, including clipping and filtering [12], probabilistic techniques [13], and coding techniques [14]. Clipping and filtering clips the OFDM signal to a predefined threshold and uses a filter to eliminate out-of-band radiation.

Probabilistic techniques use multiple candidates to represent the same information and select the one with the lowest PAR for transmission. Although coding rates can be low, coding techniques guarantee a low PAR.

1.2 Structure and Contributions of the Thesis

This thesis focuses mainly on the PAR reduction of OFDM systems. Chapter 2 describes the OFDM system and the PAR. The existing PAR-reduction techniques are discussed in Chapter 3. Chapters 4 to 6 study the mechanism of peak clipping and propose several new PAR-reduction algorithms. Chapter 7 concludes this thesis and provides suggestions for future study.

The clipping noise is analyzed in Chapter 4 as a series of parabolic pulses under tone-reservation constraints. We first consider the case in which the clipping noise consists of a single pulse, and generalize our analysis to the case of multiple pulses. Our analysis explains peak regrowth and the constant clipping noise power spectrum over the whole OFDM band. We also establish the roughly proportional relationship between the clipping noise at the end of several clipping and filtering iterations, and that generated in the first iteration. The constant of proportionality is estimated by using the level-crossing theory [15, 16].

Using the clipping noise analysis, we propose a constant-scaling algorithm and an adaptive-scaling algorithm for tone-reservation. These algorithms scale the filtered first-iteration clipping noise by a constant or adaptively-calculated factor to compensate for peaks above the threshold. The simulation results show that our proposed algorithms achieve a larger PAR reduction and lower complexity than the active-set algorithm.

Compared to the previous works [17, 18], our main contributions are as follows:

1. Our analysis is focused on the complex OFDM signal. Compared to the baseband real OFDM signal [17, 18], which has a Gaussian distribution, the complex OFDM signal has a Rayleigh distributed envelope and a complex phase, which make theoretical analysis more difficult.

2. We exploit a new model where the basic clipping pulse is approximated as a

parabolic magnitude function multiplied by a linear phase function. We derive the distribution of the phase change (in Appendices 4.7.B and 4.7.E), and prove that the phase change is small and can be omitted (in Appendix 4.7.D).

3. We prove all the conditions used in our analysis. Although these conditions are intuitive, proving them is nontrivial.

4. We extend the frequency spectrum analysis from the single pulse case ([19]) to the multiple pulses case. Our analysis explains peak regrowth and the constant clipping noise power spectrum over the whole OFDM band.

5. We propose two algorithms to find the near-optimal scaling factor. Compared to the active-set algorithm [20], our algorithms obtain a larger PAR reduction with lower complexity.

6. We propose a fast method to calculate the PAR and to find the clipping noise. We provide a necessary condition of large peaks. Only the samples that satisfy this condition need to be calculated. Because the number of such samples is small, the execution time[1] of calculating the PAR and finding the clipping noise is small. This method can also be used in other PAR-reduction techniques such as Selective Mapping (SLM) and Partial Transmit Sequences (PTS).

Chapter 5 focuses on reducing the PAR by using the sign-selection technique, which optimizes the signs of data symbols to minimize the PAR. We first propose an adaptive mapping scheme to eliminate the need for side information at the receiver. To solve the discrete optimization problem associated with the sign-selection technique, we propose in Section 5.3 three new probabilistic algorithms (the random selection, modified PTS, and recursive partial sequence methods) to find better suboptimal solutions.

Stochastic search and optimization techniques [21] can also be used to find good suboptimal solutions for the discrete optimization problem. In Section 5.4, we develop two Cross-Entropy (CE)-based PAR-reduction algorithms. Near-optimal solutions are obtained with lower complexity by simultaneously modifying the probabilities of

[1]In this thesis, the execution time is counted as the required number of real multiplications/additions.

the signs of all the subcarriers via the CE method. With a fixed number of iterations, these two algorithms obtain the same PAR reduction as the derandomization method [22, 23] with an $\mathcal{O}(N \log N)$ complexity. They also offer a flexible tradeoff between PAR reduction and execution time. The simulations show that, for the same level of PAR reduction, our algorithms require less execution time than the SLM and derandomization methods. To the best of our knowledge, our work is this first application of the CE method to PAR reduction.

Section 5.5 considers PAR reduction based on clipping-noise guided sign-selection. The key idea of clipping-noise guided sign-selection (CGS) is to iteratively flip the signs of those subcarriers with high levels of clipping noise. In each iteration, the key task is to determine the number and locations of such subcarriers. We develop suitable criteria for this task and derive two CGS algorithms that can handle both unitary (e.g., M-ary Phase Shift Keying (PSK)) and non-unitary (e.g., M-ary Quadrature amplitude modulation (QAM)) signal constellations. Although a direct comparison among PTS and SLM and CGS is not possible (as PTS and SLM use much less side information), CGS gains about 1–2 dB over these two methods for a 256-subcarrier system. A fair comparison by fixing the amount of side information is possible among CGS, derandomization and tone reservation. In this case, for a 256-subcarrier system, CGS outperforms these two methods by about 1 dB.

In Chapter 6, we generalize the Rudin-Shapiro sequences [24–26] to trade off the coding rate and the PAR. The generalized sequence is constructed from an initial PSK sequence. Its PAR is then two times that of the initial sequence, and the coding rate is increased to $\frac{\log_2 N+2m-2}{Nm}$, where N is the number of subcarriers, and m is the length of the initial sequence. We show that the PAR of the generalized sequence can be further reduced by optimizing the signs of the initial sequence.

Chapter 2

OFDM Systems

2.1 OFDM Signalling

OFDM transmits data symbols through a set of orthogonal subcarriers [27]. At the transmitter, data symbols on each subcarrier are chosen from a given M_k-point signal constellation ($k = 0, ..., N-1$), each of time duration T_s. M_k may be different for different subcarriers if bit-loading algorithms are used (e.g., in DMT systems). On the other hand, usually $M_k \equiv M$ in wireless communication systems. Each N data symbols form an OFDM block $\mathbf{X} = [X_0, ...X_{N-1}]$, which are modulated to N subcarriers and then added together and up-converted to the carrier frequency for transmission. The transmitted signal can be written as

$$x_c(t) = \Re\left[\sum_l x_l(t - lT)e^{j2\pi f_c t}\right], \tag{2.1}$$

where $\Re[x]$ represents the real part of x, f_c is the carrier frequency, and $x_l(t)$ is the l-th baseband equivalent time-domain OFDM symbol,

$$x_l(t) = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} X_{l,\langle k+N\rangle} e^{j2\pi k\Delta f t}, \quad 0 \le t \le T, \tag{2.2}$$

with $\langle k + N\rangle$ denoting $(K + N)$ modulo N, $\Delta f = 1/T$ representing the frequency spacing, and $T = NT_s$ being the time duration of the OFDM symbol. Because different OFDM symbols do not overlap, only one OFDM symbol needs to be considered, and the subscript "l" can be dropped.

Eq. (2.2) can be calculated more conveniently in the discrete-time domain. By sampling $x(t)$ at frequency $f_s = JN/T$, where J is the oversampling factor, the

discrete-time OFDM symbol x_n can be written as

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} X_{(k+N)} e^{j\frac{2\pi}{JN}nk}, \quad n = 0, ..., JN-1. \tag{2.3}$$

When $J = 1$, the above equation reduces to the Nyquist rate sampling case. Eq. (2.3) can be implemented by using a length-(JN) IFFT operation with the input vector

$$\mathbf{X}_{\text{ext}} = [X_0, ..., X_{N/2-1}, \underbrace{0, ..., 0}_{(J-1)N \text{ zeros}}, X_{N/2}, ..., X_{N-1}]. \tag{2.4}$$

Thus, $\mathbf{X}_{\text{ext}}$ is extended from $\mathbf{X}$ by using the so-called zero-insertion scheme, i.e., by inserting $(JN-1)$ zeros in the middle of $\mathbf{X}$.

In the literature, the zero-padding scheme, which appends $(JN-1)$ zeros at the end of $\mathbf{X}$, is also used. By using the zero-padding scheme, (2.2) and (2.3) may be written as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi k\Delta f t}, \quad 0 \le t \le T_s \tag{2.5}$$

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi}{JN}nk}, \quad n = 0, ..., JN-1, \tag{2.6}$$

and the implementation of (2.6) is a length-(JN) IFFT with the input vector

$$\mathbf{X}_{\text{ext}} = [X_0, X_1, ..., X_{N-1}, \underbrace{0, ..., 0}_{(J-1)N \text{ zeros}}]. \tag{2.7}$$

The difference between the zero-insertion and zero-padding schemes is the position of the carrier frequency f_c. Fig. 2.1 illustrates the frequency spectrum of both schemes. If zero-insertion is used, f_c is in the middle of the OFDM spectrum, $X_{N/2}, ..., X_{N-1}$ are modulated to the lower side-band (i.e., to the negative frequencies in the baseband equivalent model), and the $(N/2-1)$-th and $(N/2)$-th subcarriers are of the highest and lowest frequencies, respectively. On the other hand, if zero-padding is used, f_c is on the left end of the OFDM spectrum, no lower side-band exists, and the 0-th and $(N-1)$-th subcarriers are of the highest and lowest frequencies, respectively. The zero-insertion scheme matches practical situations and is easy to use for clipping analysis by using the level-crossing theory. However, these two schemes

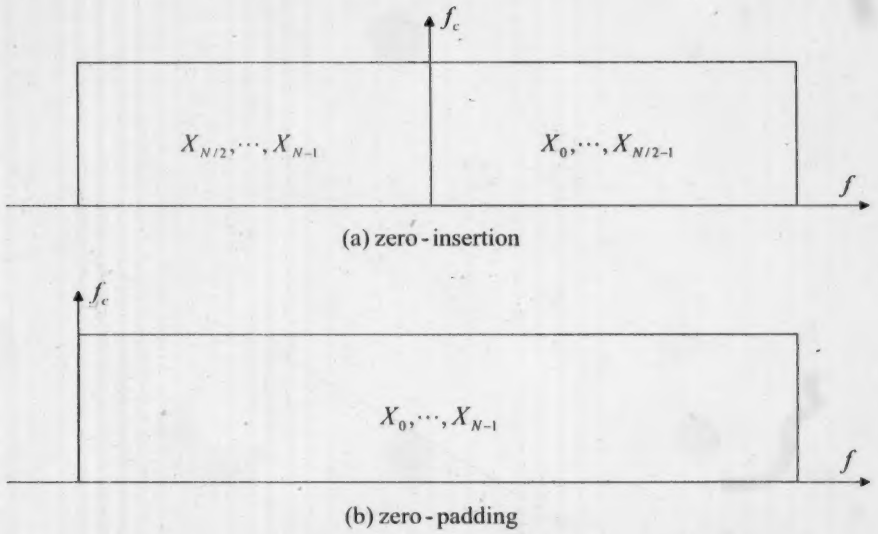


Figure 2.1: OFDM spectrum of zero-insertion and zero-padding schemes

lead to no difference in PAR statistics. In this thesis, we will alternatively use both schemes to facilitate the analysis.

At the receiver, the received signal is first down-converted to baseband and is partitioned into signal blocks $y_l(t)$, each with time duration T. Then the l-th OFDM block can be extracted from the $y_l(t)$ by using a set of orthogonal signal basis $e^{-j2\pi k\Delta f t}$, $k = -N/2, ..., N/2-1$. Also, because any pair of $y_l(t)$ will not overlap, the subscript "l" can be dropped in our analysis.

In practice, the received signal $y(t)$ is first sampled at the Nyquist frequency to obtain the discrete-time signal y_n. Then, the OFDM block is demodulated by using a length-N FFT operation in accordance with

$$Y_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y_n e^{-j\frac{2\pi}{N}nk}, \quad k = 0, ..., N-1. \tag{2.8}$$

2.2 Modulation and Demodulation Procedure

The block diagram of the OFDM system is shown in Fig. 2.2. In this figure, at the transmitter, the input bitstream is first coded by using a FEC encoder. Then the coded serial bitstream is parsed into N parallel bitstreams by using the Serial-to-Parallel (S/P) converter. Each of these parallel bitstreams is subsequently converted

to complex data symbols X_k. An IFFT converter is then used to modulate the OFDM symbols to discrete-time OFDM signals one by one. The data symbols in each OFDM block are therefore modulated into different subcarriers. After adding the cyclic prefix, the discrete-time OFDM signal is converted to a serial signal by using the Parallel-to-Serial (P/S) converter. The obtained discrete-time signal is then transferred into the continuous-time domain by using a Digital-to-Analog (D/A) converter. Finally, this signal is amplified by using an HPA and is up-converted to the carrier frequency to facilitate its transfer in the actual wireless channel.

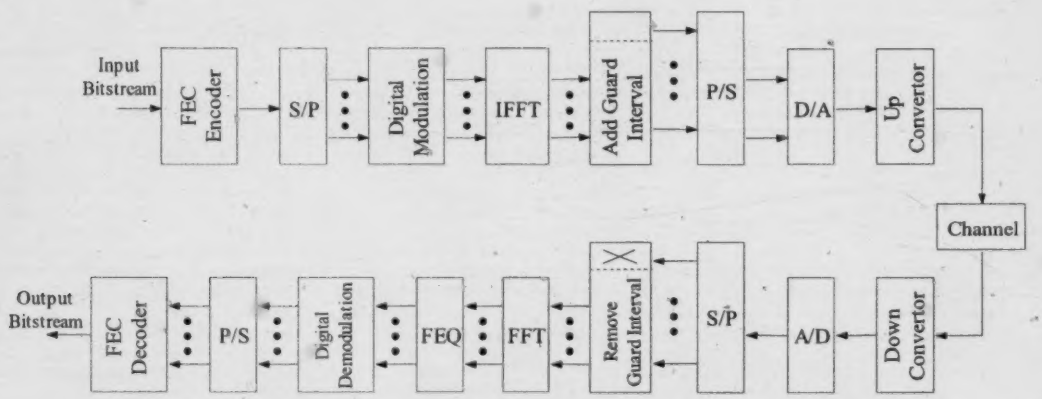


Figure 2.2: The block diagram of the OFDM system

At the receiver, the received analog bandpass signal is first down-converted to an analog baseband signal. After the Analog-to-Digital (A/D) conversion, the obtained digital signal is parsed into parallel data symbols, and the cyclic prefixes are removed. Then, the remaining data symbols are demodulated by using an FFT converter and are equalized by using a set of one-tap FEQ. The FEQ output symbols are then converted back to a serial bitstream by the digital demodulation and the P/S conversion. After FEC decoding, the input bitstream is recovered at the receiver end.

2.3 Cyclic Prefix and FEQ

The cyclic prefix (CP) of a discrete-time signal x_n is the last v samples of x_n. It is inserted at the beginning of x_n to combat the ISI without using complicated equalization techniques. Because of the multipath delay spread, signal dispersion and

overlapping will occur, leading to ISI. In other words, if the channel impulse response is h_n, $n = 0, ..., L_h - 1$, with L_h representing the length of h_n, the received signal y_n (without considering the channel noise) is the linear convolution of x_n and h_n. However, with the use of a cyclic prefix, y_n can be written as the circular convolution of x_n and h_n, provided that $v \geq L_h - 1$; i.e.,

$$y_n = x_n \otimes h_n, \tag{2.9}$$

where $\otimes$ denotes the circular convolution. In this case, after the FFT operation and dropping the cyclic prefix, we have

$$Y_k = X_k H_k, \quad k = 0, ..., N-1, \tag{2.10}$$

where H_k's are the N-point FFT of h_n. Therefore, if H_k's are known, X_k can be recovered at the receiver by using a set of FEQ $W_k = 1/H_k$ in accordance with

$$X_k = Y_k W_k = Y_k / H_k. \tag{2.11}$$

The cyclic prefix is only a replica of x_n and will not affect our PAR analysis. Hence, this prefix will not be considered in the following analysis.

2.4 Peak-to-Average Power Ratio

In OFDM systems, because the transmitted signal is the sum of a set of modulated signals, the peak power of the transmitted signal can be very high compared to its average power. Although occurring only with low probability, such large peaks have negative ramifications for the overall system. For instance, the HPA for RF transmission has to have a large linear range, which, however, is inefficiently used. Moreover, the distortion incurred by the nonlinearity of the HPA leads to in-band distortion and out-of-band radiation. The in-band distortion leads to increased BER [1]. On the other hand, the out-of-band distortion may severely interfere with the signal transmitted in the adjacent frequency bands.

The PAR of the transmitted signal can be defined as the ratio of the instantaneous power over the average power of the transmitted signal [28]:

$$\mathrm{PAR}_{x_c(t)} \triangleq \frac{\max |x_c(t)|^2}{\mathrm{E}\{|x_c(t)|^2\}}, \quad 0 \leq t \leq T, \tag{2.12}$$

where $\mathrm{E}\{\cdot\}$ represents the mean value of $(\cdot)$. On the other hand, the PAR problem can also be measured by using the baseband equivalent signal $x(t)$. Because [29]

$$\max|x_c(t)| \approx \max|x(t)|, \tag{2.13}$$

and

$$\mathrm{E}\left\{|x_c(t)|^2\right\} \approx \frac{1}{2}\mathrm{E}\left\{|x(t)|^2\right\}, \tag{2.14}$$

we have

$$\mathrm{PAR}_{x_c(t)} \approx 2\mathrm{PAR}_{x(t)}. \tag{2.15}$$

The above definition of the PAR can be called the continuous-time PAR. In practical situations, usually the PAR is calculated based on the oversampled baseband equivalent signal x_n obtained from (2.3), in accordance with

$$\mathrm{PAR} \triangleq \frac{\max|x_n|^2}{\mathrm{E}\{|x_n|^2\}}. \tag{2.16}$$

This PAR distribution is referred to as the discrete-time PAR in this thesis. It was shown in [30, 31] that Nyquist sampling ($J = 1$) may not capture all peaks of $x(t)$. Therefore, oversampling is necessary to approximate the continuous-time PAR by using the discrete-time PAR. It has been shown [28] that for an acceptable approximation, the oversampling factor J is required to be $J \geq 4$.

From (2.3) or (2.6),

$$\mathrm{E}\left\{|x_n|^2\right\} = \frac{1}{N}\sum_k |X_k|^2.$$

If X_k are independent, identically distributed (i.i.d.) random variables, the PAR of an OFDM system is al = N. Therefore, in practice, a statistical definition of PAR is more frequently used. An OFDM signal is said to have a peak at ξ with probability P_c if

$$\Pr[\mathrm{PAR}(\mathbf{X}) \leq \xi] = P_c. \tag{2.17}$$

The PAR Complementary Cumulative Distribution Function (CCDF), also called the clip probability, is defined as $P(\xi) = \Pr[\mathrm{PAR}(\mathbf{X}) > \xi] = 1 - P_c$; i.e., the probability that PAR exceeds ξ is $1 - P_c$. For example, the PAR distribution for different oversampling factors and for different N's are shown in Figures 2.3 and 2.4, respectively, which show that $J = 4$ can provide an acceptable approximation to the continuous-time PAR distribution.

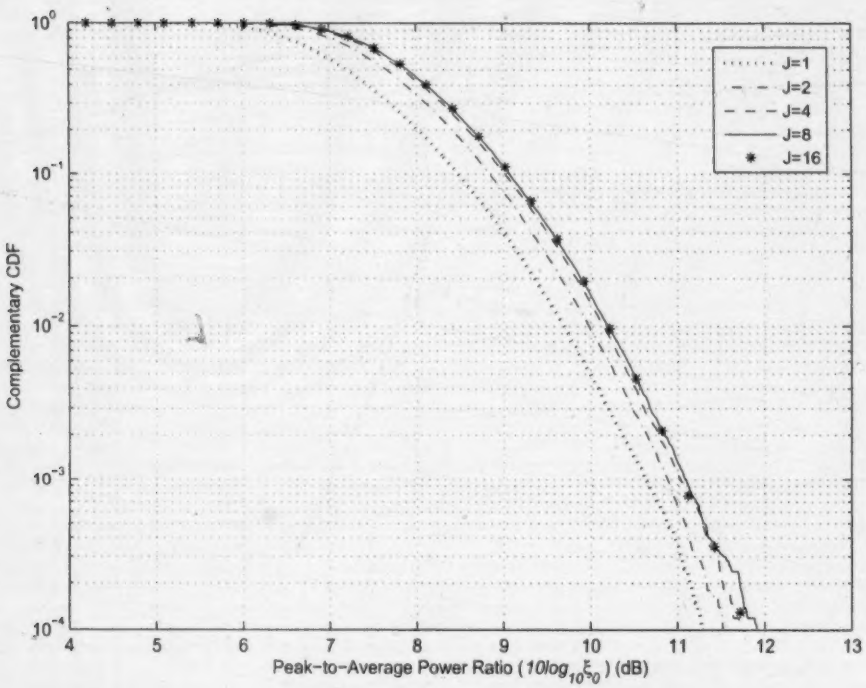


Figure 2.3: PAR distribution for different oversampling factors, $N = 128$.

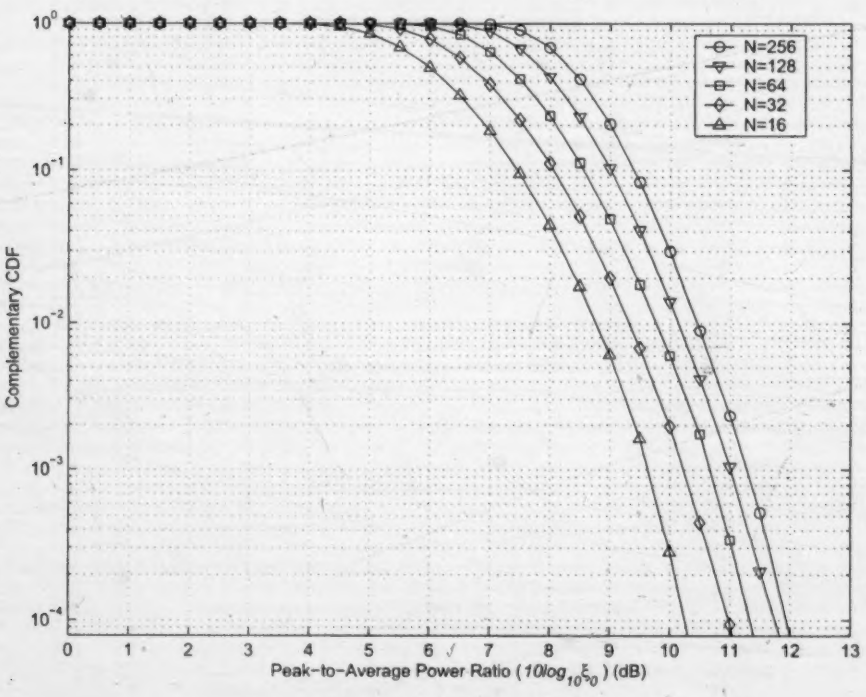


Figure 2.4: PAR distribution for different N, $L = 4$.

Chapter 3

PAR-Reduction Techniques

3.1 High-Power Amplifiers

The purpose of PAR reduction is to counteract the nonlinear effect of the HPA. Usually, HPAs are characterized as memory-less nonlinear amplifiers in accordance with

$$g(x(t)) = F(|x(t)|)e^{j(\phi(t)+\Phi(|x(t)|))}, \tag{3.1}$$

where $g(x(t))$ is the output of the HPA; $x(t) = |x(t)|e^{j\phi(t)}$ is the time domain signal input to the HPA; $F(|x(t)|)$ and $\Phi(|x(t)|)$ are, respectively, the AM/AM and the AM/PM distortion functions, where AM denotes the Amplitude Modulation, and PM denotes the Phase Modulation. Usually HPAs can be partitioned into three categories: the Soft Limiter (SL), the Solid State Power Amplifier (SSPA), and the Traveling-Wave Tube (TWT). Their characteristics can be described as follows.

3.1.1 Soft Limiter

The Soft Limiter [32] is the simplest model of the HPA. It introduces no distortion in the phase of the input signal and simply clips the signal magnitude when it exceeds a threshold. Therefore, the output of the soft limiter can be written as

$$g(x(t)) = \begin{cases} Ae^{j\phi(t)}, & |x(t)| > A, \\ x(t), & \text{otherwise}, \end{cases} \tag{3.2}$$

where $A > 0$ represents the threshold of the SL.

3.1.2 Solid State Power Amplifier

The SSPA is the most commonly used amplifier in wireless communications. The output of SSPA can be written as [1]

$$g(x(t)) = \frac{|x(t)|}{\left(1+\left(\frac{|x(t)|}{A}\right)^{2p}\right)^{\frac{1}{2p}}} e^{j\phi(t)}; \tag{3.3}$$

i.e., it introduces no distortion in the signal phase. When $p \to \infty$, the SSPA becomes the SL. Usually, $p = 3$ for a practical SSPA.

3.1.3 Traveling-Wave Tube

TWTs are wideband amplifiers widely used in satellite communications [33,34]. The AM/AM and AM/PM functions of TWT can be written as [28]

$$\begin{aligned} F(|x(t)|) &= \frac{|x(t)|}{1+\left(\frac{|x(t)|}{2A}\right)^2}, \\ \Phi(|x(t)|) &= \frac{\pi}{3}\frac{|x(t)|^2}{|x(t)|^2+4A^2}. \end{aligned} \tag{3.4}$$

As a comparison, the AM/AM functions of SL, SSPA (for $p = 3$ and $p = 10$), and TWT are shown in Fig. 3.1. The AM/PM function of TWT is shown in Fig. 3.2.

3.1.4 PAR Distribution and BER Performance

When X_k are PSK symbols, an upper bound of the PAR can be easily obtained as [35]

$$\xi \leq 1 + \frac{2}{N}\sum_{n=1}^{N-1} |R_X(n)|, \tag{3.5}$$

where $R_X(n)$ is the aperiodic autocorrelation function of X_k defined as

$$R_X(n) = \sum_{k=0}^{N-n-1} X_{k+n} X_k^*, \tag{3.6}$$

with $(\cdot)^*$ representing the complex conjugate. With the assumption that X_k are i.i.d. random variables and based on the central limit theory, x_n can be approximated

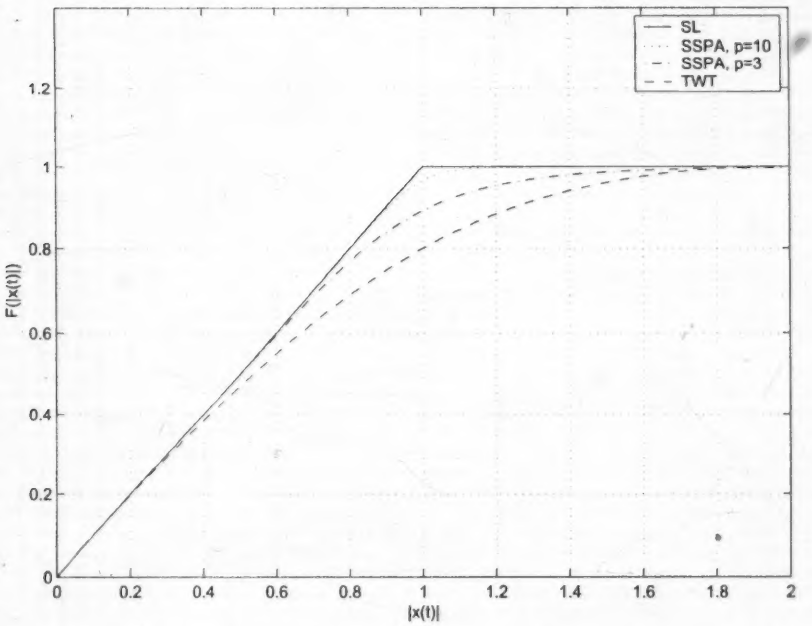


Figure 3.1: AM/AM functions of SL, SSPA (for $p = 3$ and $p = 10$), and TWT.

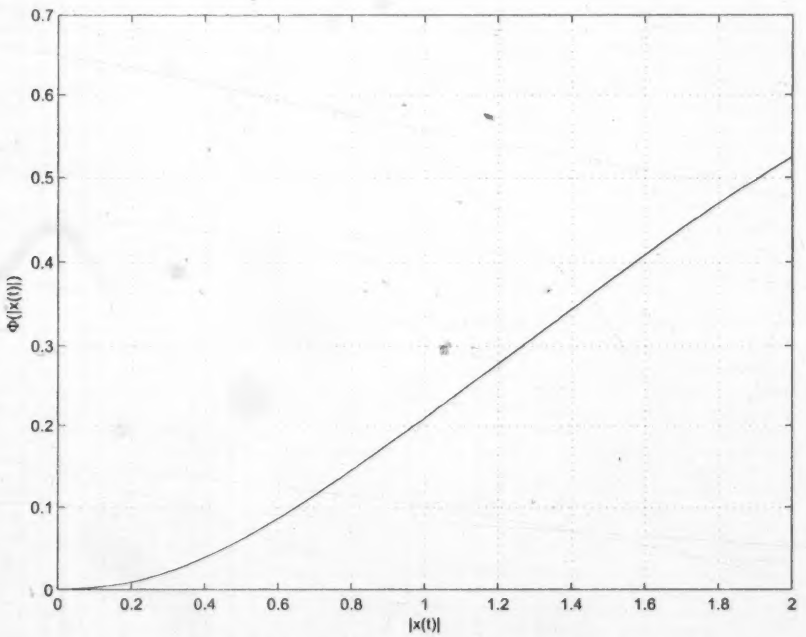


Figure 3.2: AM/PM function of TWT.

as (complex) Gaussian random variables when N is large. Then, $|x_n|$ is Rayleigh distributed. The PAR distribution can be approximated as [36]

$$\Pr[\xi \leq \xi_p] \approx \left(1 - e^{-\xi_p/2\sigma^2}\right)^{\alpha N}, \tag{3.7}$$

where σ^2 is the variance of the real or imaginary part of x_n, and $\alpha = 2.8$ (obtained from empirical experiments). More accurate approximations are also available [37–56]. In [46], by using the theory of the level-crossing rate and normalizing $r(t) = |x(t)|$ such that $2\sigma^2 = 1$, the probability that all peaks are lower than r is

$$\Pr(\max[r(t)] < r) \approx \begin{cases} \left(1 - \frac{re^{-r^2}}{\bar{r}e^{-\bar{r}^2}}\right)^{\sqrt{\frac{\pi}{3}}N\bar{r}e^{-\bar{r}^2}}, & \text{for } r > \bar{r}, \\ 0, & \text{for } r \leq \bar{r}, \end{cases} \tag{3.8}$$

where $\bar{r}$ is empirically obtained as $\bar{r} = \sqrt{\pi}$ for Quadrature Phase Shift Keying (QPSK) and slightly lower for 16QAM. The PAR distribution can then be found by replacing r and $\bar{r}$ with $\sqrt{\xi_p}$ and $\sqrt{\xi_p}$, respectively.

The effect of signal clipping on BER performance has been extensively studied. Such analysis focuses mainly on the Signal-to-Noise-plus-Distortion Ratio (SNDR) and BER after the passage of $x(t)$ through an SL.

Clipping $x(t)$ by the SL introduces a clipping noise $f(t) = x(t) - g(x(t))$, which includes in-band distortion and out-of-band radiation. Fig. 3.3 illustrates the power spectral density (PSD) of unclipped and clipped OFDM signals. 9 dB clipping leads to relatively small (−51 dB) out-of-band radiation. However, deeper clipping, e.g., 6 dB and 3 dB clipping, significantly increases out-of-band radiation to −31 dB and −21 dB, which may be unacceptable in practical communications. In [45], it is shown that by applying Bussgang's theorem [32], the clipped signal can be written as

$$\hat{x}_n = \alpha x_n + d_n, \quad n = 0, ..., JN - 1, \tag{3.9}$$

where x_n and d_n are uncorrelated, and the attenuation factor α can be found as

$$\alpha = 1 - e^{-\gamma^2} + \frac{\sqrt{\pi}\gamma}{2}\text{erfc}\,(\gamma) \tag{3.10}$$

with $\gamma = A/\sqrt{P_i}$.

For large N and small A, the clipping noise can be approximated as a Gaussian process. For Nyquist-rate clipping (x_n is Nyquist-rate sampled), no out-of-band radiation exists. In this case, the SNDR for Additive White Gaussian Noise (AWGN)

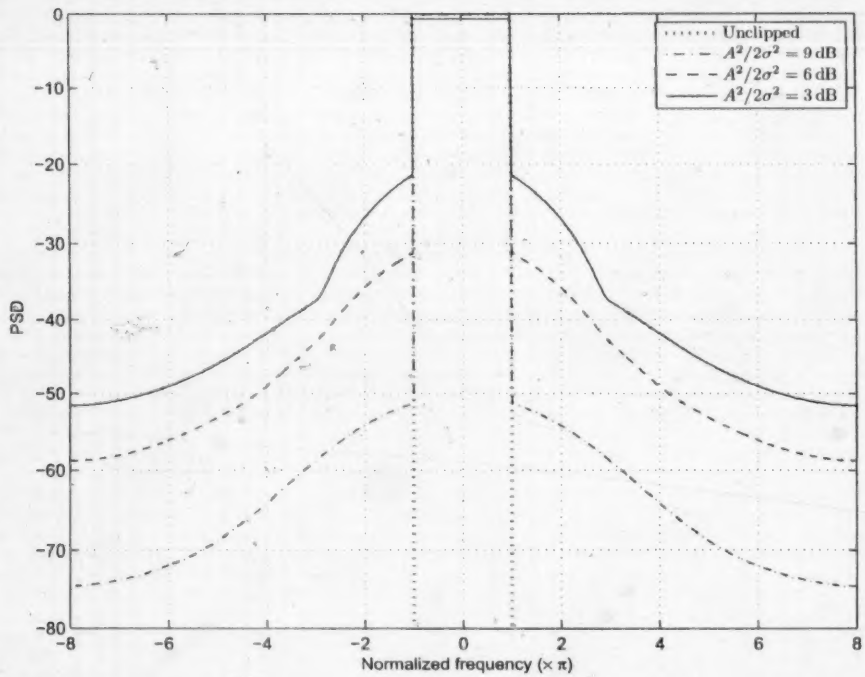


Figure 3.3: Power spectral density of unclipped and time-domain-clipped OFDM signals.

channel is given by [45]

$$\text{SNDR} = \frac{K_\gamma E_s/N_0}{(1-K_\gamma)E_s/N_0 + 1}, \tag{3.11}$$

where E_s/N_0 is the ratio of the signal energy over the noise Power Spectral Density (PSD) after clipping, and $K_\gamma = \alpha^2/(1-e^{-\gamma^2})$. The BER of QPSK is then given by

$$P_b = Q(\text{SNDR}), \tag{3.12}$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2} dt$.

The assumption of Gaussian clipping noise holds only for small A. When A is large, the clipping noise is a series of pulses which can be approximated as parabolic arcs [15]. Based on this approximation, the BER of a real-valued time domain OFDM signal (which is used for DMT applications) can be calculated as [17]

$$P_b = \frac{8N(L-1)}{\sqrt{3}L} e^{-\gamma^2/2} Q\left(\left[\frac{3\pi\gamma^2}{\sqrt{8(L^2-1)}}\right]^{1/3}\right), \tag{3.13}$$

where a square constellation of L^2 points with a minimum distance of $2d$ is assumed.

In [49], the performance of OFDM with a strictly limited peak-power requirement is analyzed. The analytical results show that the clipping technique exhibits the lowest Input Back-off (IBO) requirement compared to the probabilistic techniques, especially when N is large.

3.2 PAR-Reduction Techniques

Various techniques have been proposed to reduce the PAR, including clipping-based techniques, probabilistic techniques and coding techniques.

3.3 Categorization of the PAR Problem

PAR research can be divided into two categories: PAR distribution and BER evaluation, and the development of efficient PAR-reduction techniques. Fig. 3.4 illustrates this categorization. PAR-reduction techniques can be further divided into two categories: techniques with distortion and without distortion. Techniques with distortion are usually based on signal clipping and lead to continuous solutions where the OFDM signals are modified in a continuous manner. These techniques include iterative clipping and filtering [12, 19, 57, 58], and companding [59–81]. A detailed discussion of iterative clipping and filtering will be presented in the next subsection.

For techniques with distortion, distortion-cancellation techniques are necessary in order to reduce a possible BER loss. A task common to both categories is the development of low complexity algorithms.

Distortion-less PAR-reduction techniques are more attractive because they do not increase the BER. In fact, the BER can even be decreased by exploiting the inherent redundancy. These techniques can be divided into three categories: (1) techniques having continuous solutions such as tone reservation and Active Constellation Extension (ACE), (2) techniques having discrete solutions such as SLM, PTS and tone injection, and (3) coding techniques. Distortion-less techniques may need side information for correct detection. Developing efficient decoding techniques is also a related research area.

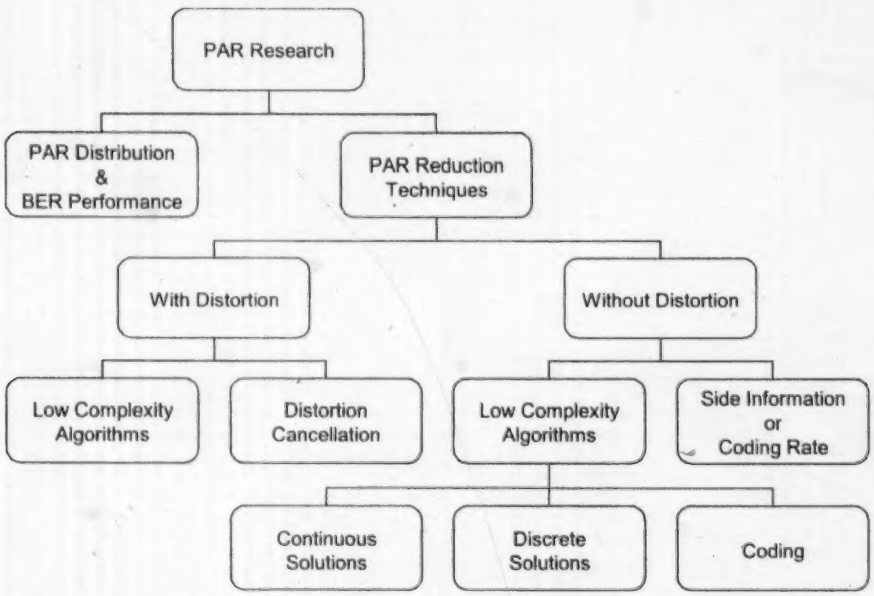


Figure 3.4: Categorization of the PAR problem

While the PAR problem has obvious practical value, it is also related to many theoretical research areas such as discrete optimization and coding theories.

3.3.1 PAR-Reduction Techniques with Distortion

The simplest PAR-reduction technique is clipping and filtering [12, 19, 57, 58]. This technique clips the OFDM signal to a predefined threshold and uses a filter to eliminate the out-of-band radiation. The purpose of this technique is to satisfy the spectral constraints so that the OFDM signal will not interfere with communications in the neighboring frequency bands. The inband distortion, however, cannot be eliminated, leading to increased BER. Because the probability of large peaks is small, the inband distortion is also small when the clipping threshold is large. Consequently, the BER increase may be small when, e.g., 4QAM is used, and may be tolerable in some applications. On the other hand, the inband distortion may be limited to a predefined strength [82–85] to reduce the BER increase, with the cost of degraded PAR-reduction performance. In [36], each clipping noise sample is multiplied by a window function (e.g., Gaussian, Kaiser, or Hamming) to suppress the out-of-band noise. Because the overall effect is the convolution of the clipping samples and the windowing function,

this approach leads to a bandwidth increase. Alternatively, the inband distortion can also be partly canceled at the receiver with additional computational cost. [86–89].

Filtering out-of-band radiation can be done in the time domain by using a lowpass filter [12] or in the frequency domain by using an FFT/IFFT pair [19, 58, 90]. By using frequency domain filtering, which requires less execution time than using a time-domain lowpass filter, the clipping noise is converted to the frequency domain by using oversampled FFT, and all out-of-band terms are set to zero. An IFFT is also required to convert the filtered clipping noise back to the time domain. Nevertheless, a side-effect of filtering is peak regrowth, shown in Fig. 3.5. After filtering, the signal peaks grows higher than the clipping threshold (but lower than the original peaks). Generally, peak regrowth can be combatted by iterative clipping and filtering [90]. However, the convergence is extremely slow after several iterations. To speed up the convergence, [90] proposes using a slightly lower clipping threshold, say, 95% of the desired PAR level [57], to better suppress peak regrowth.

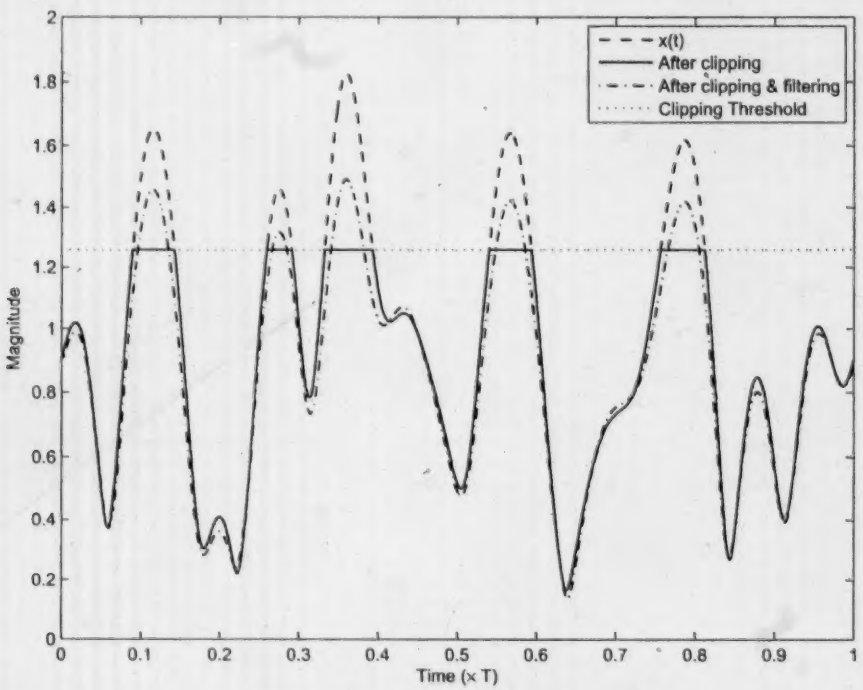


Figure 3.5: Peak regrowth.

Clipping the OFDM signal is equivalent to passing the signal through a Soft Lim-

iter. The clipped portion of the signal cannot be perfectly recovered. In contrast, the companding transform [59–81] passes the OFDM signal through a smooth nonlinear function

$$y(t) = g(x(t)),$$

where $x(t)$ is the input OFDM signal, and $g(x)$ is the nonlinear companding transform, to "compress" the large peaks in $x(t)$. The probability that $y(t)$ has large peaks is then much smaller than before companding transforming.

At the receiver, $x(t)$ is recovered by using $g^{-1}(t)$, the inverse transform of $g(t)$. The recovered OFDM signal can be written as

$$\hat{x}(t) = g^{-1}(g(x(t)) + z(t)),$$

where $z(t)$ is the channel noise. The companding transform has two drawbacks. First, the signal bandwidth is expanded after the transform.[1] Second, the channel noise $z(t)$ is enhanced at the receiver when recovering the compressed samples. The bandwidth expansion and noise enhancement depend on the target PAR level of the companding transform. If the target PAR level is low, filtering is required at the transmitter to eliminate the out-of-band radiation, which introduces inband distortion and peak regrowth. At the receiver, the BER may be unacceptable.

3.3.1.1 Clipping Noise Cancelation

Clipping noise can be partly cancelled at the receiver. In [91], the clipping noise is estimated and then cancelled by using oversampled signal reconstruction, which reconstructs the clipped samples by interpolating the oversampled signal. This method requires the transmission of a portion of the out-of-band radiation to the receiver. Thus, the OFDM bandwidth must be expanded by 25% or more. [92] proposes a decision-aided reconstruction method to mitigate the clipping noise. After channel equalization, this method first makes a hard decision (in the frequency domain) on the received signal samples $\tilde{x}_n$ to estimate the data symbols, and then converts them back to the time domain to obtain $\tilde{x}_n$. If the channel distortion and channel noise are

[1] The bandwidth of the compressed signal is the same as the original signal only when the companding transform is applied to the Nyquist-rate samples. However, the PAR-reduction performance is poor in this case.

omitted[2], $\hat{x}_n = x_n$ at the positions of $|x_n| \leq A$, where A is the clipping threshold. On the other hand, $\tilde{x}_n$ contains samples larger than A at the positions of $|x_n| > A$, which may be a better estimation of x_n than $\hat{x}_n$ if the clipping is not severe. Therefore, x_n can be estimated more accurately by

$$\bar{x}_n = \begin{cases} \hat{x}_n, & \hat{x}_n \leq A, \\ \tilde{x}_n, & \text{otherwise.} \end{cases}$$

This method is further improved in [86–89] by reconstructing the clipping noise, and is later applied to coded OFDM [93,94] and MIMO OFDM [95]. By using (3.9), the received samples after the FFT operation can be written as

$$Y_k = \alpha H_k X_k + H_k D_k + Z_k, \tag{3.14}$$

where Y_k, H_k, D_k and Z_k are the received signal, channel response, clipping noise, and AWGN on the k-th subcarrier, respectively. Assuming that H_k is known, a coarse estimation $\bar{X}_k$ can be obtained by making a hard decision on Y_k/H_k. By converting $\bar{X}_k$ to the time domain, clipping it in the same fashion as at the transmitter, and converting the clipped signal back to the frequency domain, we have

$$\tilde{X}_k = \alpha \bar{X}_k + \bar{D}_k. \tag{3.15}$$

Assume that most $\bar{X}_k$ are correct. Then $\bar{D}_k \approx D_k$, and one can use $\bar{D}_k$ to obtain a better estimation of X_k; i.e.,

$$\hat{Y}_k = Y_k - H_k \bar{D}_k = \alpha H_k X_k + H_k (D_k - \bar{D}_k) + Z_k. \tag{3.16}$$

This procedure can be repeated to improve the estimation accuracy.

3.3.2 Tone-Reservation

The tone-reservation technique reserves N_r tones for PAR reduction and uses the remaining $(N - N_r)$ tones for data transmission [18, 20, 96–104]. The reserved tones may be randomly selected, or be selected from the subcarriers that have low SNR and are not suitable for data transmission. The tone-reservation ratio $R = \dfrac{N_r}{N}$ is typically

[2]The channel distortion and channel noise degrade the performance of clipping noise cancelation.

small. The peak-canceling signal $c(t)$ is generated based on the reserved tones, and the peak-reduced signal is given by

$$\hat{x}(t) = x(t) + c(t) = \frac{1}{\sqrt{N}} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} (X_k + C_k) e^{j2\pi kt/T}, \quad 0 \le t \le T,$$

where $\mathbf{C} = [C_{-\frac{N}{2}}, ..., C_{\frac{N}{2}-1}]$ is the set of peak-canceling tones. $\hat{x}(t)$ is amplified by the HPA and transmitted to the receiver. Denote $\mathcal{C}$ as the signal space of all possible $\mathbf{C}$ vectors.

Let $\mathcal{R} = \{i_0, ..., i_{N_r-1}\}$ be the locations of the reserved tones, where $-\frac{N}{2} \le i_0 < i_1 < \cdots < i_{N_r-1} \le \frac{N}{2} - 1$. Let the index set $\mathcal{R}^c$ be the complement of $\mathcal{R}$ in $\mathcal{N} = \{-\frac{N}{2}, ..., \frac{N}{2} - 1\}$. The constraint on $c(t)$ is that $\mathbf{C}$ must satisfy $C_k \equiv 0$ for $k \in \mathcal{R}^c$. On the other hand, $\mathbf{X}$ must satisfy $X_k \equiv 0$ for $k \in \mathcal{R}$. $\mathbf{X}$ and $\mathbf{C}$ are not allowed to be nonzero on the same subcarriers; i.e.,

$$X_k + C_k = \begin{cases} X_k & k \in \mathcal{R}^c, \\ C_k & k \in \mathcal{R}. \end{cases} \tag{3.17}$$

Clearly, this technique reduces the normalized system throughput to $(1 - R)$. For a frequency-selective fading channel (ignoring the nonlinear amplification), demodulation is done on a per-tone basis. Thus, (3.17) allows the reserved tones to be readily discarded. With this method, the BER of the data tones is the same as that of the original OFDM system. However, the BER of the whole system is slightly increased due to the slightly increased average transmit power.

With the tone-reservation, the PAR is redefined as

$$\xi = \frac{\max |x(t) + c(t)|^2}{E\{|x(t)|^2\}}; \tag{3.18}$$

that is, the peak-canceling signal $c(t)$ is excluded from the calculation of the average power to prevent the solution of a $c(t)$ having an averaged power much larger than that of $x(t)$. The denominator in Eq. (3.18) is a constant. Thus, $\mathbf{C}$ must be chosen to minimize the maximum of the time-domain signal:

$$\mathbf{C}^{(\text{opt})} = \arg \min_{\mathbf{C} \in \mathcal{C}} \max_{0 \le t \le T} \left| \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} (X_k + C_k) e^{j2\pi kt/T} \right|^2. \tag{3.19}$$

Eq. (3.19) can be reformulated as a Quadratically Constrained Quadratic Program (QCQP) [18]:

$$\begin{aligned} &\min_{\mathbf{C}\in\mathcal{C}} E \\ &\text{subject to } |x_n + \mathbf{q}_n\mathbf{C}|^2 \leq E \end{aligned} \tag{3.20}$$

for $n = 0, 1, ..., JN-1$, where $\mathbf{q}_n$ is n-th row of the IDFT matrix. (3.20) is convex, and the global optimum exists [105]. However, finding the optimal solution requires a high computational cost. Suboptimal solutions are typically employed.

The simplest optimization technique for tone-reservation is iterative clipping and filtering [90]. In each iteration, this technique clips the OFDM signal to a predefined threshold A. The clipped signal is then filtered such that the clipping noise exists only on reserved tones. The convergence rate of this technique is extremely slow [18].

The controlled clipper algorithm [18] iteratively calculates the peak-reduced time-domain OFDM signal as follows:

$$\bar{\mathbf{x}}^{(i+1)} = \bar{\mathbf{x}}^{(i)} - \mu \underbrace{\sum_{|\bar{x}_n^{(i)}|>A} \alpha_n^{(i)} \mathbf{P}_n^{\|_2}}_{\mathbf{c}}, \tag{3.21}$$

where $\bar{\mathbf{x}}^{(i)}$ is the peak-reduced signal at the i-th iteration, $\bar{\mathbf{x}}^{(0)} = \mathbf{x}$, $\mathbf{x}$ is the OFDM signal, A is the target magnitude upper bound of the peak-reduced signal, $\alpha_n^{(i)}$ are the clipping noise samples at the i-th iteration, $\mathbf{P}_n^{\|_2}$ is the prototype peak-canceling signal, and $\mathbf{c}$ is the sample vector of the peak-canceling signal $c(t)$. The convergence rate of this algorithm slows down after several iterations, and many iterations are usually required to obtain a reasonable PAR reduction [18].

An active-set algorithm is proposed in [20] for tone-reservation. This algorithm first approximates the peak boundary (a circle centered at the origin of the complex plane) as a polygon. For example, Fig 3.6 shows an octagon of radius A. The magnitude of point X can be approximated by

$$\begin{aligned} |X| \approx |X|_{\text{Approx}} &= \max\left[\Re[X], \Re\left[Xe^{j\pi/4}\right], \Re\left[Xe^{j\pi/2}\right], \Re\left[Xe^{j3\pi/4}\right]\right] \\ &= \max\left[\Re[X], \Im[X], \Re\left[Xe^{j\pi/4}\right], \Im\left[Xe^{j\pi/4}\right]\right], \end{aligned}$$

where $\Re[x]$ and $\Im[x]$ represent the real and imaginary parts of x, respectively. On the other hand, all points that satisfy $|X|_{\text{Approx}} \leq A$ are within the octagon of radix

A. In other words, if we reduce the approximated peak of an OFDM signal to A, its actual peak would be only slightly larger than A.

With this polygonal approximation, the complex OFDM vector $\mathbf{x}$ is written as a real vector $\hat{\mathbf{x}}$ consisting of the real and imaginary parts of $\mathbf{x}$ and its phase-shifted

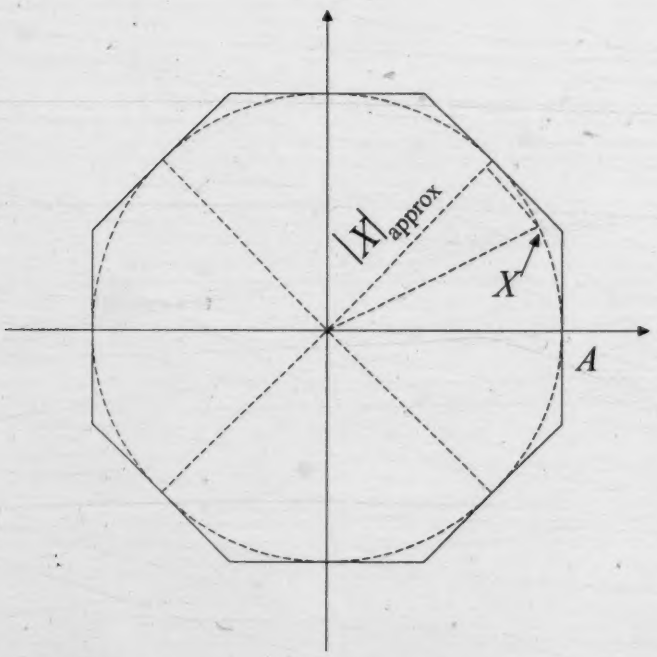


Figure 3.6: Polygonal approximation of the peak boundary.

The active-set algorithm uses the same prototype peak-canceling signal $\mathbf{P}_n^{\|_2}$ (but is rewritten to a real vector $\hat{\mathbf{p}}$, as explained above) to reduce the PAR. This algorithm maintains an active set containing the peaks of $\hat{\mathbf{x}}$, whose magnitudes are reduced to the same level as that in previous iterations. Each sample x_{n_i} in the active set is associated with a peak-canceling kernel $\hat{\mathbf{p}}_i$ (a shifted version of $\hat{\mathbf{p}}$ whose peak is at n_i). These $\hat{\mathbf{p}}_i$ are weighted and summed together to form the peak-canceling signal $\bar{\mathbf{p}}$. In each iteration, the weighting factors of $\hat{\mathbf{p}}_i$ are calculated by solving a set of l linear equations, where l is the iteration number, to find a proper optimization

direction. Then, at least all peaks of both $\hat{\mathbf{x}}$ and $\bar{\mathbf{p}}$ outside the active set[3] are tested to find a proper optimization step size μ. The peak-canceling signal $\bar{\mathbf{p}}$ is weighted by μ and is subtracted from $\hat{\mathbf{x}}$. Then, the magnitudes of all samples in the active set are equally reduced. The largest peak outside the active set is also reduced to the same magnitude of the samples in the active set; and is included in the active set. After several iterations, the PAR is reduced to a moderate level.

3.3.3 Active Constellation Extension

The ACE technique [106, 107] allows the constellation be extended (by the clipping noise) so that the minimum Euclidian distance between any two constellation points does not increase. For example, the shaded areas in Fig. 3.7 are the feasible extension regions for the 16QAM constellation.

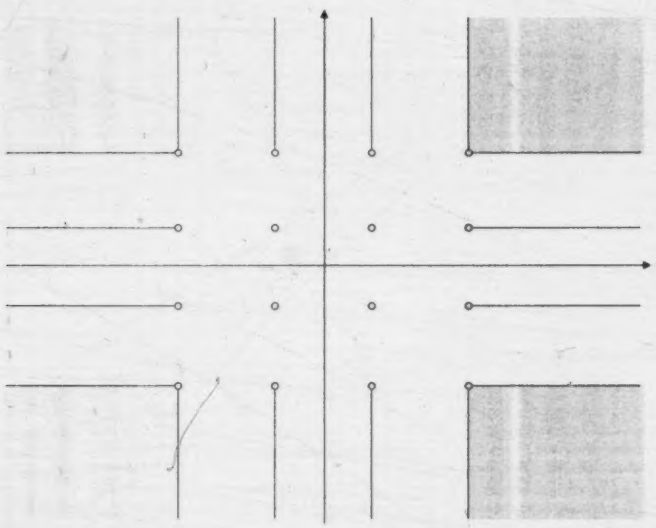


Figure 3.7: Feasible extension region for 16-QAM constellation.

Finding the optimal peak-canceling signal is similar to (3.20) but with the additional constraint that the peak-canceling signal in the frequency domain does not allow the use of any value beyond the feasible region. The optimization methods discussed above can be applied to the ACE technique.

The ACE technique does not reduce the throughput. However, this technique slightly increases the average transmit power, leading to a slightly increased BER. Its

[3]To ensure the active set criteria, generally all samples of $\hat{\mathbf{x}}$ outside the active set must be tested.

PAR-reduction capability depends on the signal constellation. Only the outer signal points in the constellation are allowed to move. If PSK or 4QAM constellations are used, all input data symbols are free to move in the feasible region, and a large PAR reduction is obtained. On the other hand, when an MQAM constellation, $M > 4$, is used, in average only $\frac{4\sqrt{M}-4}{M}N$ subcarriers in each OFDM block are available for PAR reduction. Thus, the PAR-reduction capability of ACE is small for large QAM constellations.

3.3.4 Probabilistic Techniques

By modifying the phase, amplitude and/or subcarrier position of input symbols, these techniques use several candidate OFDM signals to represent the same information, and select the one with the lowest PAR. Side information may be required at the receiver for correct detection.

3.3.4.1 Phase-Adjustment Techniques

A widely used technique is the modification of the phase of the input modulation symbols to reduce the PAR [108–111]. Let $\mathbf{X} = [X_0, ..., X_{N-1}]$ be an OFDM vector, and let $\mathbf{s} = [s_0, ..., s_{N-1}] = [e^{j\phi_0}, ..., e^{j\phi_{N-1}}]$ be a phase-adjustment vector. Then, the objective function is

$$\min_{\mathbf{s}} \left(\text{PAR of IDFT}(\mathbf{s} \odot \mathbf{X})\right), \tag{3.23}$$

where $\odot$ represents the element-wise multiplication. The optimal solution $\mathbf{s}^{(\text{opt})}$ may be sent to the receiver as side information for the correct detection of the input modulation symbols. The distribution of s_k, $k = 0, ..., N-1$, determines the minimum PAR that can be obtained. Assume that s_k are i.i.d., N is large, and $\mathbf{s}^{(\text{opt})}$ is selected from K randomly generated phase-adjustment vectors $\{\mathbf{s}^{(1)}, ..., \mathbf{s}^{(K)}\}$ according to a predefined distribution of s_k. If the PAR is calculated on the Nyquist-rate discrete-time-domain samples (i.e., $J = 1$ for calculating the IDFT), it has been proved that the PAR of IDFT($\mathbf{s}^{(\text{opt})} \odot \mathbf{X}$) is minimized when $\mathrm{E}\{s_k\} = 0$ [112]. Thus, we may choose s_k from 1 or -1 with equal probability to minimize the resulting throughput loss. On the other hand, if the PAR is calculated on the oversampled discrete-time-domain samples (i.e., $J > 1$) or on the continuous-time-domain signal, choosing s_k

from a larger set (e.g., $s_k \in \{1, j, -1, -j\}$) leads to only minor improvement in the PAR reduction [113]. In this thesis, we use $s_k \in \{1, -1\}$.

Eq. (3.23) describes a combinatorial optimization problem over the N dimensional binary space $\{1, -1\}^N$. Let $s_0 \equiv 1$ without loss of generality. The size of the search space is 2^{N-1}, which grows exponentially with N. Also, finding the PAR of a phase sequence requires the evaluation of all the JN samples of the phase-adjusted OFDM block. The execution time of finding the optimal phase sequence is thus prohibitively large when N is large. Many suboptimal techniques have been proposed to reduce the optimization complexity by reducing the search space and/or by efficiently computing the PAR of each phase sequence.

[114] proposes an method to estimate the PAR of a phase sequence without computing all the JN samples. If the estimated PAR is large, the phase sequence is rejected. Otherwise, the exact PAR is calculated as usual and compared with that of other sequences to find the optimum sequence. In [115], the polygonal approximation of the peak boundary [20] (3.22) is used to reduce the number of multiplications in finding the PAR of a sequence, at the cost of an increased number of additions. However, because the total number of arithmetic operations (multiplications and additions) is still large, this method does not reduce the execution time when the multiplier-accumulator [116] is used[4]. In [104], we propose a fast algorithm to compute the PAR. (This algorithm will be discussed in Section 4.3.) The above methods can be combined with other methods that reduce the search space to further speed up the optimization.

To reduce the search space, [117] proposes two criteria for constructing a phase sequence set that may lead to a low PAR. By studying the PAR relationship between two sign sequences, [117] proves that the upper bound of the PAR difference between two sign sequences is statistically maximized when the two sign sequences are orthogonal. Moreover, the element-wise product of any two sequences in the set should not be periodical or similar to periodical. However, a systematic construction method has not been found.

The SLM method [118–120] uses a set of K, $K \ll 2^{N-1}$, randomly generated phase

[4]By using the multiplier-accumulator, the execution time of the multiplication is comparable to that of the addition.

sequences $\{\mathbf{s}^{(1)}, ..., \mathbf{s}^{(K)}\}$. For each OFDM block, the phase sequence $\mathbf{s}^{(i)}$ leading to the lowest PAR is selected to adjust the phases of the OFDM block. Thus, the search space is reduced to a set of K phase sequences. SLM reduces the probability of large peaks. If the probability that $\mathbf{s}_i$ $(i = 1, ..., K)$ leads to a PAR larger than ξ is P_ξ, then the probability that all $\mathbf{s}_i$ leading to a PAR larger than ξ is $P_\xi^K < P_\xi$. SLM requires K IFFTs for each OFDM block, and the minimum side information is $\log_2 K$ bits.

The PTS technique [13,121–124] partitions each OFDM block $\mathbf{X}$ into K, $K \ll N$, disjoint subblocks $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_K]$. The sign sequence $\mathbf{s}$ is also partitioned into K corresponding subblocks $\mathbf{s} = [\mathbf{s}_1, ..., \mathbf{s}_K]$, where the elements within each $\mathbf{s}_i$ are the same; i.e., $\mathbf{s}_i = [s_i, s_i, ..., s_i]$. The size of the search space is then reduced from 2^{N-1} to 2^{K-1}, making an exhaustive search possible. 2^{K-1} IFFTs are required for each OFDM block if the PAR of each sign sequence is calculated by using an IFFT. However,

$$\mathbf{x} = \text{IFFT}[\mathbf{s} \odot \mathbf{X}] = \sum_{k=1}^{K} s_k \text{IFFT}[\mathbf{X}_k], \tag{3.24}$$

where $\mathbf{x}$ is the vector of time domain signal. Thus, IFFT$[\mathbf{X}_k]$ can be calculated and stored before searching for the suitable phase sequence. Then, optimizing the PAR of each OFDM block requires only K IFFTs and $2^{K-1}JN(K-1)$ complex additions. The minimum side information is $(K-1)$ bits (because $\mathbf{s}_1 \equiv 1$).

It has been proved that global optimal $\mathbf{s}^{(\text{opt})}$ leads to a constant PAR [22]. By using the derandomization method, the PAR can be iteratively reduced to less than $c \log N$ where c is a constant [23,125]. This approach outperforms other probabilistic techniques, but with a complexity of $\mathcal{O}(N^2)$.

A sign-flipping method is proposed to reduce the execution time of PTS [126], where, in each iteration, the sign of a subblock is flipped between $+1$ and -1, and the one leading to the lower PAR is retained. The whole search space can be formulated as a binary tree. The sign-flipping method searches only one branch of the tree. By searching more branches, a larger PAR reduction can be obtained with increased execution time [127]; for example, we may search only the all-1 phase sequence and its neighbors having a Hamming distance less than or equal to r [128]. In addition to the 2^{K-1} phase sequences (i.e., K subblocks) that modify the OFDM block to reduce the PAR, extra modifications of the OFDM block can be obtained by complex

conjugating, frequency-reversing or circular-shifting a subblock, or multiplying a subblock by a predefined phase sequence[5] [129]. A dual-layer search scheme is proposed in [130], where an OFDM block is divided into D subgroups, called divisions, and each division is further divided into M subblocks. A suboptimal sign sequence can be found by, e.g., first optimizing the signs of each subblock and then optimizing the signs of each division. [131] proposes using a sphere decoder [132–135] to find the optimal phase sequence for PTS. In [136], suboptimal solutions are found by using a trellis search.

In [137, 138], the IFFT algorithm (used for converting the OFDM block to the time domain) is modified such that the original OFDM block is first processed by the IFFT algorithm for k stages (the complete N-point IFFT algorithm has $\log_2 N$ stages) to obtain a intermediate sequence. Different phase sequences are applied to the intermediate sequence and then converted to the time domain by completing the last $n - k$ stages. Because the last $n - k$ stages involve only small-sized IFFTs, the total execution time is then reduced. By exploiting the structure of IFFT, a set of specially designed phase sequences can be constructed such that, after calculating the PAR of a "base" phase sequence, the PAR of the other phase sequences can be calculated from this "base" sequence by using additions only [139–141].

A threshold may be used to reduce the execution time. That is, we search the whole solution space until a phase sequence leading to a PAR lower than a threshold is found. The average execution time is low. However, the latency varies because, with a small occurrence probability, some OFDM blocks may need to test a large number of phase sequences to reach the threshold. Such a latency problem can be alleviated by using an input buffer and an output buffer [142].

An alternative to the PTS technique is based on the quantization of the continu-

[5] If the IFFT of a subblock $\mathbf{X}_k$ has been obtained, the IFFT of these operations can be readily obtained without any arithmetic operation.

ously valued (sub)optimal solution. Rewrite (3.24) as

$$\mathbf{x} = \underbrace{\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,M} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,M} \\ \vdots & \vdots & & \vdots \\ A_{JN,1} & A_{JN,2} & \cdots & A_{JN,M} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_M \end{bmatrix}}_{\mathbf{s}}, \tag{3.25}$$

where $[A_{1,m}, A_{2,m}, \cdots, A_{JN,m}]^T$ is the IFFT of $\mathbf{X}_m$. For each row of $\mathbf{A}$, an optimal $\mathbf{s}$ can be found either by sorting the elements in this row of $\mathbf{A}$ according to the decreasing order of their magnitudes and letting s_k alternatively be the phase and negative phase of the corresponding $A_{n,m}$ [143], or by projecting a predefined vector to the null space of this row of $\mathbf{A}$ [144]. Then, the optimal continuous valued $\mathbf{s}$ is quantized to discrete values. Eventually, JN candidates of $\mathbf{s}$ are found, and the one leading to the lowest PAR is selected.

3.3.4.2 The Side Information Issue

Side information may be embedded in the OFDM block for transmission, or may be eliminated at the cost of lowered throughput or coding gain. Side information must be embedded in the OFDM block before PAR optimization because, otherwise, peak regrowth may occur [145]. [146] proposes using a marking algorithm to embed side information for PTS. For every possible sign sequence, this algorithm rotates every other data symbol in the subblocks having negative signs by $\pi/4$. A sign sequence with the lowest PAR is then searched. This method does not lead to any throughput loss. However, if a high-order constellation is used, this method requires a large computational cost for decoding, but the decoding results may not be reliable [147]. In [28], side information is inserted at the beginning and the middle of the OFDM block if SLM or interleaving (discussed later) is used, or is inserted into the first subblock of the next OFDM block if PTS is used.

[148] proposes an SLM scheme without transmitting side information for coded OFDM. This scheme encodes data symbols by using a (n, k) non-binary code. Only the parity-check symbols are allow to change signs by using one of the K predefined sign sequences. The receiver also knows the K sign sequences. At the receiver, a

new set of parity-check symbols are constructed by encoding the received k data symbols, and then compared with the received parity-check symbols to estimate the sign sequence. This estimated sequence is then compared to the K sign sequences to find the one used at the transmitter. This scheme has no throughput loss due to side information, but the coding gain is lowered.

[122] shows that transmitting side information is not necessary if the Hamming distance of the phase sequences used in PTS or SLM is large, and the phase-adjusted symbols do not fall on any modulation constellation points. By using a simplified maximum likelihood (ML) decoder at the receiver, data symbols are recovered with a slightly increased BER.

When a low PAR is required, the amount of side information is large, and finding a set of phase sequences to satisfy the condition in [122] might be difficult. In this case, one may modify the modulation constellation to eliminate side information. In [149], we propose an adaptive mapping scheme to map the modulation constellation $\mathcal{Q}_M$ to $\mathcal{Q}_{2M}$. We will discuss this scheme in Chapter 5. In [150,151], hexagonal constellations are used to avoid side information. Fig. 3.8 illustrates an example of the 91-point hexagonal (91-Hex) constellation. The 64 points marked by $\circ$ and $\triangle$ are used to carry six information bits. Thus, the 91-Hex constellation has the same throughput and the same minimum Euclidian distance as the square 64QAM constellation. Its average symbol energy is $10.36d^2$, where d is the minimum Euclidian distance, which is slightly smaller than that of the square 64QAM ($10.50d^2$).

The signs of 27 outer points $\triangle$ can be modified (between $\triangle$ and $\times$) to minimize the PAR. Therefore, no side information is required at the receiver. However, on average, only $\bar{N}_s = 27$ subcarriers are allowed to change for PAR reduction, and the average search space is reduced to $S = (1 + \frac{\bar{N}_s}{N})^N$.

3.3.4.3 Amplitude/Phase-Adjustment Techniques

The PAR of OFDM signals can be reduced by modifying the amplitude and phase of the data symbols. The tone injection technique [18] expands the constellation. For example, Fig. 3.9 shows the extension of 16QAM, where a point A_1 of the original 16QAM (the shaded area) can be mapped to A_1, A_2, A_3 or A_4. A search in a discrete solution space is then required to find the optimum mapping. To simplify the search,

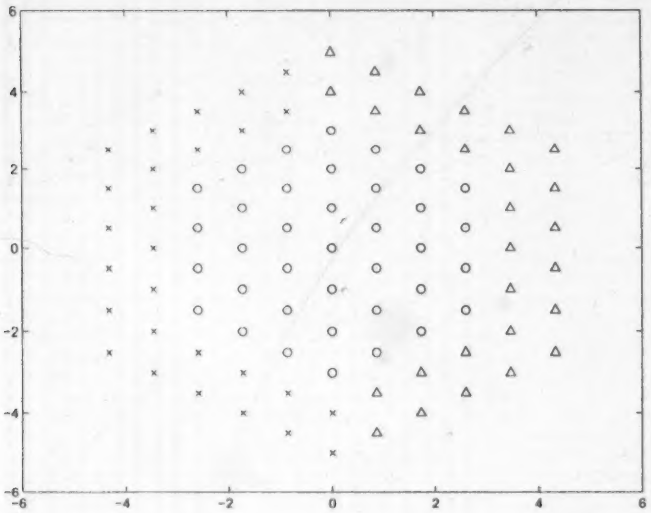


Figure 3.8: The 91-point hexagonal constellation.

usually only the points within the dashed square are used, and each 16QAM symbol has two mapping choices. No side information is required at the receiver. However, the average power is slightly increased.

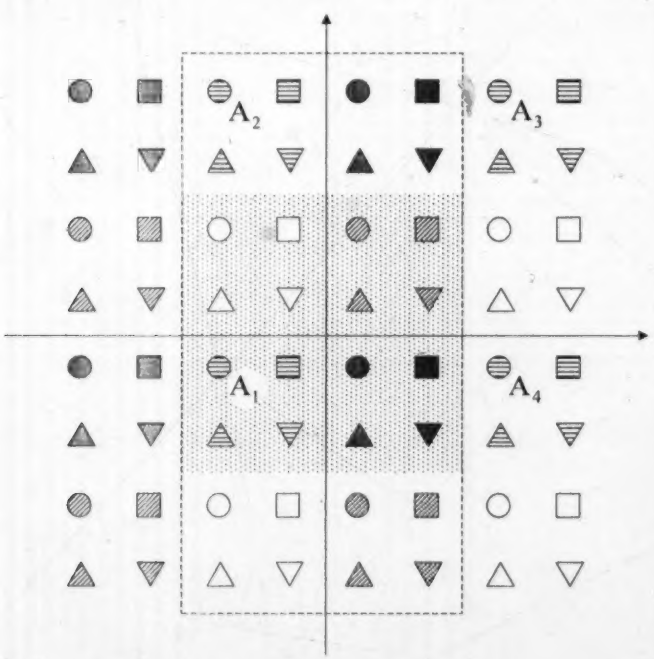


Figure 3.9: Tone Injection.

Trellis shaping, which is ordinarily used for reducing the average power at the same bit rate [152], can also be used to reduce the PAR [153–156]. This technique uses the same constellation expansion as tone injection. However, by using a convolutional code, this technique gives rise to higher throughput than tone injection. For example, trellis shaping encodes each five bits to a point in Fig. 3.9. Similar to tone injection, the trellis shaping technique first maps four bits to a point, say A_1, in the shaded area. The other bit s is extended to a 2-bit inverse syndrome of a rate-1/2 convolutional codeword $\mathbf{z} = s(\mathbf{H}^{-1})^T$, where $\mathbf{H}$ is the generator matrix of the convolutional code, and $(\cdot)^T$ represents the matrix transpose. By using the Viterbi decoder, a codeword $\mathbf{y}$ is generated such that $\mathbf{y} + \mathbf{z}$ selects the one among $A_1, ..., A_4$ that leads to the lowest PAR.

3.3.4.4 Scrambling and interleaving

Data permutation can also be used to randomly generate K independent candidates for PAR reduction [157–160]. The permutation can be performed either bit-wise or symbol-wise. In [161], a selective scrambling approach is proposed where 4 scramblers are used for picking up the candidate with the lowest PAR, and 2 bits (00,01,10, or 11) are concatenated with information bits to indicate which scrambler is used. By using a convolutional encoder [162] or a shift-register with a feedback branch [163], side information can be embedded into the scrambled vector as the scrambler's initial state. Similar to the method proposed in [163], a guided scrambling method is proposed in [164] for SLM and PTS, where the OFDM block (or subblocks if PTS is used) is first augmented by different pattern labels (binary vectors), and then divided by the scrambling polynomial. Side information (i.e., the label of the selected pattern) is then embedded in the scrambled vector.

Usually, L pilot tones are inserted in each OFDM block to acquire the channel state information. As long as the pilot tones are equal-power and equally spaced, changing the position of the pilot tone set does not affect channel estimation [165]. Thus, one may reduce the PAR of an OFDM block by shifting the pilot tone set to the optimal position [166]. Because the average power of the pilot tones is larger than that of the data tones, the former can be easily separated from the latter without side information at the receiver. PAR can also be reduced by setting some carefully-

selected subcarriers to zero to obtain a zero-padded PSK [9]. This approach can also be viewed as modifying the signal constellation such that zero is included in the constellation.

3.3.5 Coding Techniques

Coding techniques transmit only the low-PAR codewords [30, 167, 168]. Golay complementary sets may be used to generate polyphase sequences with a low PAR [14].

K sequences $\mathbf{X}_1, ..., \mathbf{X}_K$ are said to be complementary if the sum of their aperiodic autocorrelation functions satisfies

$$\sum_{k=1}^{K} R_{\mathbf{X}_k}(n) = \delta(n) \sum_{k=1}^{K} R_{\mathbf{X}_k}(0), \tag{3.26}$$

where $\delta(n)$ is the Kronecker delta function, and $R_{\mathbf{X}_k}(n)$ is defined in (3.6). Golay complementary sets become Golay complementary pairs when $K = 2$. We can show that

$$\mathrm{PAR}_{\mathbf{X}_k} \leq K, \quad k = 1, ..., K. \tag{3.27}$$

The Rudin-Shapiro sequence [14] is a special case of the Golay complementary pair, which can be generated recursively as follows:

$$p_{n+1} = p_n + e^{j\theta_n} z^{2^n} q_n, \tag{3.28}$$

$$q_{n+1} = p_n - e^{j\theta_n} z^{2^n} q_n, \tag{3.29}$$

where $p_0 = q_0 = 1$, $z = e^{j\omega_0 t}$, and $e^{j\theta_n}$ are PSK symbols. The PAR of p_n or q_n is no larger than 2.

Golay complementary sequences can also be generated by concatenating or interleaving two short complementary sequences [168]. [14] shows that most Golay complementary sets are related to first-order Reed-Muller codes. Therefore, a Golay complementary sequence can be generated by

$$\mathbf{X} = [u_0, u_1, ..., u_m, c_1, c_2, ..., c_K]\mathbf{G}, \tag{3.30}$$

where $u_i \in \{0, ..., M-1\}$ $(i = 0, ..., m$, and M is an even number) are phase indices of M-PSK symbols, $c_k \in \{0, M/2\}$ $(k = 1, ..., K$, and $K = \binom{m}{2})$ defines the second-order coset, $\mathbf{G}$ is the generator matrix of the second-order Reed-Muller code with

elements of 0 or 1 and the dimension of $(m + K + 1) \times 2^m$, and the addition and multiplication operations in the matrix multiplication are defined over the modulo-M. The corresponding M-PSK OFDM block is $e^{j\mathbf{X}/M}$.

For a given c_k, the minimum Hamming distance is $d_{\min} = 2^{m-1}$, and the coding rate is

$$R = \frac{m+1}{2^m} = \frac{\log_2 N + 1}{N}, \tag{3.31}$$

where N is the length of the coded codewords. With N increasing, R goes to zero.

Many choices of c_k lead to codewords with a PAR of less than 2. If all these codewords are used, the minimum Hamming distance is $d_{\min} = 2^{m-2}$, and the increased coding rate is bounded as

$$R \leq \frac{m + 1 + \binom{m}{2}}{2^m}. \tag{3.32}$$

However, R is close to zero when N is very large.

Reed-Muller codes can be written as a boolean function $f(x_1, ..., x_m)$, where $[x_m, ..., x_1]^T$ forms a m by 2^m matrix with columns, from left to right, being the binary representation of $1, 2, ..., 2^m$, respectively. The second-order Reed-Muller code contains only the terms of x_i and $x_i x_j$ ($\forall i$ and j). [14] shows that, for any permutation π of the symbols $\{1, 2, ..., m\}$ and for any u, $u_k \in \mathbb{Z}_{2^h}$, where h is an integer,

$$a(x_1, ..., x_m) = \sum_{k=1}^{m} u_k x_k + 2^{h-1} \sum_{k=1}^{m-1} x_{\pi(k)} x_{\pi(k+1)} + u, \tag{3.33}$$

is a Golay complementary sequence over $\mathbb{Z}_{2^h}$ of length 2^m. The second term determines the value of c_k in (3.30). This equation gives $m!/2$ cosets of first-order Reed-Muller codes. Therefore, the coding rate is

$$R = \frac{m + 1 + \lfloor \log_2(m!/2) \rfloor}{2^m}. \tag{3.34}$$

[169] proves that (3.33) forms a path on a graph $G(Q)$ with vertices of $x_1, x_2, ...x_m$. If deleting k vertices of the graph results in a path, then all codewords of the coset $Q + \mathrm{RM}_q(1, m)$ (q is an even number) have a PAR no larger than 2^{k+1}. Therefore, a tradeoff is allowed between the coding rate and the PAR. Similarly, [170] proposed multiple shift codes, which also make a tradeoff between the coding rate and the PAR. The main property of multiple shift codes is

$$R_{\mathbf{X}}(n) + R_{\mathbf{Y}}(n) = 0, \quad \text{for } 1 \leq n \leq N - 1 \text{ and } n \bmod L = 0, \tag{3.35}$$

where $L \in \{1, 2, ..., N-1\}$. The PAR of $\mathbf{X}$ or $\mathbf{Y}$ is then no larger than L.

Ordinary Golay complementary sequences are restricted to PSK modulation. Recently, methods of constructing complementary sequences on high-order QAM constellations have been proposed [171, 172]. These methods use two 4QAM on the complex plane. By properly choosing the offsets of the two 4QAM, complementary sequences can be constructed on 16QAM or 64QAM. For more detail, see [172] and the references therein.

3.3.6 PAR Reduction for MIMO OFDM Systems

Interest has been growing in the application of OFDM in multiple antenna systems and the use of Space-Time Block Codes (STBCs) or Space-Frequency Block Codes (SFBCs) to combat fading and reduce the outage probability. The Multiple-Input Multiple-Output (MIMO) OFDM system also suffers from a high PAR. Generally, PAR-reduction techniques for conventional Single-Input Single-Output (SISO) OFDM systems can be directly applied to MIMO OFDM systems. Some modifications exploiting the structure of MIMO systems are also proposed in the literature [173–178].

Instead of optimizing each antenna separately as in SISO OFDM, most modifications for PAR reduction in MIMO OFDM focus on optimization over all the antennas to reduce the amount of side information and/or the computational cost (with a slight loss of PAR-reduction performance). For example, when SLM is used, each phase-adjustment vector is multiplied to all antennas. The phase-adjustment vector leading to the lowest PAR on all antennas is selected [175]. The cross-antenna rotation and inversion (CARI) method also adjusts the phase of the data symbols and swaps data symbols between two antennas to reduce the PAR [178]. When SFBCs is used in MIMO OFDM, CARI must be modified such that the SFBCs code structure is not violated [173].

Chapter 4

PAR Reduction Using Clipping-based Techniques

In this chapter, we analyze the clipping noise under the tone-reservation constraints and propose two clipping-based algorithms for peak reduction in OFDM systems [104]. To facilitate our analysis, we use the zero-inserting OFDM system where the carrier frequency is in the middle of the OFDM frequency band. The time-domain OFDM symbol $x(t)$ and its discrete-time samples x_n may be written as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X_k e^{j2\pi kt/T}, \quad 0 \le t \le T, \tag{4.1}$$

where N data symbols X_k form an OFDM block $\mathbf{X} = [X_{-\frac{N}{2}}, ..., X_{\frac{N}{2}-1}]$, and T is the OFDM symbol period, and

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} X_k e^{j2\pi \frac{nk}{JN}}, \quad n = 0, ..., JN-1, \tag{4.2}$$

where J is the oversampling factor.

4.1 Problem Formulation

The tone-reservation technique [18] reserves N_r tones for PAR reduction and uses the remaining $(N - N_r)$ tones for data transmission. The tone-reservation ratio $R = \frac{N_r}{N}$ is typically small. The simplest method to generate the peak-canceling signal is iterative clipping and filtering [90]. In each iteration, this technique clips the OFDM signal to

a predefined threshold A. The clipped signal is then filtered such that the clipping noise exists on the reserved tones only. The convergence rate of this technique is slow.

Alternatively, we scale the filtered clipping noise to form the peak-canceling signal $c(t)$. For clipping $x(t)$ using a soft limiter [32], the clipped OFDM signal $\tilde{x}(t)$ becomes

$$\tilde{x}(t) = \begin{cases} Ae^{j\theta(t)}, & |x(t)| > A, \\ x(t), & |x(t)| \leq A, \end{cases} \tag{4.3}$$

where A is the predefined threshold, and $\theta(t)$ is the phase of $x(t)$. The clipping noise is

$$f(t) = x(t) - \tilde{x}(t). \tag{4.4}$$

The clipping noise $f(t)$ consists of the segments of $x(t)$ where $|x(t)|$ exceeds A. Unless A is small, $f(t)$ is thus a series of pulses,

$$f(t) = \sum_{i=1}^{N_p} f_i(t),$$

where $f_i(t)$ is the i-th clipping pulse with pulse duration τ_i, with its amplitude maximum at t_i, and N_p is the number of clipping pulses.

The filtered clipping noise $\hat{f}(t)$ is obtained by passing $f(t)$ through a filter whose passbands are on the reserved tones. The peak-canceling signal is a scaled version of the filtered clipping noise:

$$c(t) = -\beta \hat{f}(t),$$

where β is the scaling factor to be optimized. One of our objectives is to optimize β such that the PAR is minimized. Thus, the optimization problem is

$$\min_{\beta} \max_{0 \leq t \leq T} \left| x(t) - \beta \hat{f}(t) \right|^2. \tag{4.5}$$

We will start with the analysis of clipping and filtering, and reveal the mechanism of peak regrowth; i.e., the clipped peaks may grow and exceed A after filtering. The analysis of the peak regrowth will facilitate the optimization of β. We will also give an analytical explanation of the flat spectrum of the clipping noise, which has been observed in [99] by simulation.

4.2 Analysis of Clipping and Filtering

4.2.1 Time Domain Analysis of Clipping Noise

In our analysis, we assume that the real and imaginary parts of input data symbol X_k are independent, identically distributed (i.i.d.) random variables with zero mean and variance σ^2. We also assume that A and N are large, T is small, and the OFDM bandwidth $W = N/T$ is a constant.

In DMT systems, $x(t)$ is real. Based on the central limit theory, $x(t)$ is a Gaussian random process when N is large. The clipping noise $f(t)$ is then the consequence of the upward level crossing of $x(t)$ at level A and the down level crossing of $x(t)$ at level $-A$. The level crossing of a Gaussian process has been extensively studied [15,32,179–188]. The spectrum of the clipped Gaussian process is given in [189] and the BER of the clipped DMT signal is given in [17].

In OFDM systems, $x(t)$ is a complex signal. Let $x(t) = x_R(t) + jx_I(t) = r(t)e^{j\theta(t)}$, where $x_R(t)$, $x_I(t)$, $r(t) \geq 0$ and $\theta(t)$ are the real and imaginary parts, the magnitude and the phase of $x(t)$, respectively. Based on the central limit theory, $x_R(t)$ and $x_I(t)$ are i.i.d. Gaussian random processes with zero mean and variance[1] σ^2, $r(t)$ is a Rayleigh process, $\theta(t)$ is uniformly distributed between $[0, 2\pi)$, and $r(t)$ is independent to $\theta(t)$. The power of $x(t)$ (i.e., $r^2(t)$), is a χ^2 process with two degrees of freedom.

The clipping noise $f(t)$ is the consequence of the upward level crossing of $r(t)$ at level A, or equivalently, of $r^2(t)$ at level A^2. The level crossing of a χ^2 process has been studied in [190, 191]. We will use the results from these studies to analyze the time and frequency domain characteristics of the clipping noise $f(t)$.

The level crossing rate (the expected number of crossings of level A per second) can be found as [16]

$$\lambda_A = \frac{\dot{\sigma}}{\sqrt{2\pi}} \frac{A}{\sigma^2} e^{-A^2/2\sigma^2}, \tag{4.6}$$

where [15]

$$\dot{\sigma}^2 = E\{\dot{x}_R^2(t)\} = E\{\dot{x}_I^2(t)\} = \frac{1}{2\pi} \int \omega^2 S(\omega) d\omega,$$

and $S(\omega)$ is the PSD of $x_R(t)$ or $x_I(t)$. When N is large, $S(\omega)$ is (approximately)

[1] If $N_r = RN$ tones are reserved, the variance of $x_R(t)$ and $x_I(t)$ is then $\sigma^2 R$.

constant over a fixed frequency band $[-W/2, W/2]$. Then, we have

$$\dot{\sigma}^2 = \frac{(\pi N)^2 \sigma^2}{3T^2} = \frac{\pi^2}{3} W^2 \sigma^2. \tag{4.7}$$

By substituting (4.7) into (4.6), the level crossing rate is

$$\lambda_A = \sqrt{\frac{\pi}{6}} \frac{A}{\sigma} \frac{N}{T} e^{-A^2/2\sigma^2}. \tag{4.8}$$

With our assumption of a large A, each up-crossing of level A leads to a clipping pulse. Therefore, the average number of clipping pulses in one OFDM signal duration can be calculated as

$$\bar{N}_p = E\{N_p\} = \lambda_A T = N \sqrt{\frac{\pi}{6}} \frac{A}{\sigma} e^{-A^2/2\sigma^2}. \tag{4.9}$$

The clipping pulse duration τ is a Rayleigh random variable with a probability density function [190]

$$p(\tau) = \frac{\pi \tau}{2\bar{\tau}^2} \exp\left(-\frac{\pi \tau^2}{4\bar{\tau}^2}\right), \tag{4.10}$$

where $\bar{\tau}$ is the mean of τ. Because $\lambda_A \bar{\tau} = \Pr[r(t) > A]$, $\bar{\tau}$ can be calculated as

$$\bar{\tau} = \frac{\Pr[r(t) > A]}{\lambda_A} = \frac{\sigma^2 \sqrt{2\pi}}{\dot{\sigma} A} = \sqrt{\frac{6}{\pi}} \frac{\sigma}{AW}. \tag{4.11}$$

Let us consider a clipping pulse $f_i(t)$ that reaches its maximum magnitude at t_i and has a time duration τ_i. That is, $f_i(t) = (r(t) - A)e^{j\theta(t)}$ within its pulse duration and is zero elsewhere. Eqs. (4.10) and (4.11) imply that, most probably, τ is very small in practical OFDM systems. Then, $r(t)$ can be approximated as a parabolic function by using its Taylor's series expansion at $t = t_i$. Let $\Delta t_i = t - t_i$. Because $r(t_i) > A$, $\dot{r}(t_i) = 0$ and $\ddot{r}(t_i) < 0$, we have

$$\begin{aligned} r(t) = r(t_i + \Delta t_i) &\approx r(t_i) + \dot{r}(t_i)\Delta t_i + \frac{1}{2}\ddot{r}(t_i)\Delta t_i^2 \\ &= r(t_i) + \frac{1}{2}\ddot{r}(t_i)\Delta t_i^2. \end{aligned} \tag{4.12}$$

With this approximation, $r(t_i + \Delta t_i)$ is symmetric to t_i. Then, $r(t_i - \tau_i/2) \approx r(t_i + \tau_i/2) \approx A$, and

$$\tau_i \approx \sqrt{-\frac{8(r(t_i) - A)}{\ddot{r}(t_i)}}. \tag{4.13}$$

Let $b_i = -\ddot{r}(t_i)$. We have

$$r(t_i + \Delta t) - A \approx -\frac{1}{2} b_i \Delta t_i^2 + \frac{1}{8} b_i \tau_i^2, \quad -\frac{\tau_i}{2} \le \Delta t_i < \frac{\tau_i}{2}. \tag{4.14}$$

Now let us look at the phase $\theta(t) = \theta(t_i + \Delta t_i)$. The phase change within $-\tau_i/2 \leq \Delta t_i \leq \tau_i/2$ is generally small. The phase of $f_i(t_i + \Delta t_i)$ is determined by all the constituent frequency components of $x(t)$, where $x(t)$ is a band-limited signal. The phase change of the k-th frequency component, from $t = t_i$ to $t = t_i + \frac{\tau_i}{2}$, is

$$\Delta\theta_k = 2\pi k \frac{\tau_i}{2T}, \quad k = -\frac{N}{2}, -\frac{N}{2}+1, ...\frac{N}{2}-1.$$

Substituting τ_i by $\bar{\tau}$, we find

$$\Delta\theta_k = \frac{\sqrt{6\pi}k\sigma}{NA}, \quad k = -\frac{N}{2}, -\frac{N}{2}+1, ...\frac{N}{2}-1.$$

The largest phase change happens on $k = -\frac{N}{2}$, and its value does not depend on N. By letting $\theta_k = \theta_{\frac{N}{2}}$ for all k, the phase variation of $f_i(t)$ from $t = t_i$ to $t = t_i + \frac{\tau_i}{2}$ is upper-bounded by $\frac{\sqrt{6\pi}\sigma}{2A}$. Clearly, the upper bound is quite loose, and the actual phase variation of $f_i(t)$ is much smaller than this bound because some negative and positive phase changes may cancel each other. Nevertheless, because this upper bound is small when A is large, we can approximate $\theta(t_i + \Delta t_i) = \arcsin\left(\frac{x_I(t_i + \Delta t_i)}{r(t_i + \Delta t_i)}\right)$ by its Taylor's series expansion at $t = t_i$:

$$\theta(t_i + \Delta t_i) \approx \theta_i + \gamma_i \Delta t_i,$$

where $\theta_i = \theta(t_i)$ and

$$\gamma_i = \frac{\dot{x}_I(t_i)}{|x_R(t_i)|}.$$

Then,

$$\begin{aligned} f_i(t) = f_i(t_i + \Delta t_i) &= (r(t_i + \Delta t_i) - A)e^{j\theta(t_i + \Delta t_i)} \\ &\approx (-\frac{1}{2}b_i\Delta t_i^2 + \frac{1}{8}b_i\tau_i^2)e^{j(\theta_i + \gamma_i\Delta t_i)}, \quad -\frac{\tau_i}{2} \leq \Delta t_i < \frac{\tau_i}{2}. \end{aligned}$$

The absolute value of the phase term $\gamma_i \Delta t_i$ is most probably small and can be omitted. Appendix 4.7.B gives the conditional probability density function (pdf) and moments of γ_i given $\dot{r}(t_i) = 0$ and $r(t_i) \geq A$. Using these results, we have

$$E\{|\gamma_i|\}E\{\tau_i\} = \frac{\sqrt{2\pi}\sigma}{A}\text{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right)e^{A^2/2\sigma^2}, \tag{4.15}$$

where

$$\text{erfc}\,(x) = 1 - \text{erf}\,(x) = 1 - \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt.$$

In Appendix 4.7.D, we show that γ_i and τ_i are uncorrelated. Thus, (4.15) may provide us some information about how small $\gamma_i\tau_i$ usually is. For example, $E\{|\gamma_i|\}E\{\tau_i\} \approx 0.07\pi$ when $A/\sqrt{2}\sigma = 6\,\text{dB}$, and is 0.04π when $A/\sqrt{2}\sigma = 9\,\text{dB}$. A rigid justification of small $|\gamma_i\Delta t_i|$ requires the joint cumulative distribution function (cdf) of $\gamma_i\tau_i$, which, unfortunately, is difficult to write in the closed form. However, an upper bound for the phase change of $f_i(t)$ can be found.

Because $|f_i(t)|$ is close to zero when $|\Delta t_i|$ is close to $\tau_i/2$, we may look at the phase change within the 6 dB width of $f_i(t)$, which is defined as the time duration when $|f_i(t)|$ is no less than half of its maximum magnitude and is equal to $\tau_i/\sqrt{2}$. By using the Chebyshev inequality, we have

$$\Pr[|\gamma_i\tau_i/\sqrt{2}| \geq \delta] \leq \frac{\sigma_0^2}{\delta^2},$$

where σ_0^2 is the variance of $\gamma_i\tau_i/\sqrt{2}$, and $\delta > 0$. However, by using the Cauchy-Schwarz inequality,

$$\sigma_0^2 = E\{\frac{1}{2}\gamma_i^2\tau_i^2\} \leq \frac{1}{2}\sqrt{E\{\gamma_i^4\}E\{\tau_i^4\}}.$$

Denote the right-hand side of this inequality as σ_1^2. It is calculated by using the results of Appendix 4.7.B as

$$\sigma_1^2 = \frac{2\sqrt{6}\sigma^3}{A^3}\sqrt{2 - \mathrm{E}_1\left(\frac{A^2}{2\sigma^2}\right)\frac{A^2}{\sigma^2}e^{A^2/2\sigma^2}},$$

where

$$\mathrm{E}_1(x) = \int_1^\infty \frac{e^{-tx}}{t}dt.$$

Then, we have

$$\Pr[|\gamma_i\tau_i/\sqrt{2}| \geq \delta\sigma_1] \leq \frac{1}{\delta^2}.$$

For example, by letting $\delta = 3$, the probability that the phase change within the 6 dB width of $f_i(t)$ is larger than 0.12π for $A/\sqrt{2}\sigma = 6\,\text{dB}$, or 0.03π for $A/\sqrt{2}\sigma = 9\,\text{dB}$, is less than or equal to $1/9$. Therefore, we can omit the phase term $\gamma_i\tau_i$ and approximate the clipping pulse as a constant phase parabolic function:

$$f_i(t) = f_i(t_i + \Delta t_i) \approx (-\frac{1}{2}b_i\Delta t_i^2 + \frac{1}{8}b_i\tau_i^2)e^{j\theta_i}, \quad -\frac{\tau_i}{2} \leq \Delta t_i < \frac{\tau_i}{2}. \tag{4.16}$$

Remark 4.1. In our approximation, we implicitly assume that $f_i(t)$ has only one local maximum (at $t = t_i$). In other words, $\ddot{r}(t_i)$ is always negative. Appendix 4.7.C shows that $\Pr[\ddot{r}(t_i) > 0 | \dot{r}(t_i) = 0, r(t_i) \geq A] \to 0$ when $A \to \infty$, and in practical OFDM systems, $\Pr[\ddot{r}(t_i) > 0 | \dot{r}(t_i) = 0, r(t_i) \geq A] \approx 0$ unless A is very small. On the other hand, some papers (e.g., [17]) approximate $f_i(t)$ by expanding $r(t)$ at $t = t_i - \tau_i/2$, where $r(t_i - \tau_i/2) = A$ and $\dot{r}(t_i - \tau_i/2) \geq 0$. Then, τ_i becomes

$$\tau_i \approx -\frac{2\dot{r}(t_i - \tau_i/2)}{\ddot{r}(t_i - \tau_i/2)}$$

with an assumption that $\ddot{r}(t_i - \tau_i/2) < 0$. Although such an assumptions holds for $A \to \infty$, simulation results show that it is frequently violated even for $A = 6\,\text{dB}$.

4.2.2 Frequency Domain Analysis of Clipping Noise

The frequency spectrum of the Nyquist-rate sampled discrete-time real clipping noise is given in [17]. For the continuous-time complex clipping noise, the frequency spectrum of $f_i(t)$ is the Fourier transform of (4.16); i.e.,

$$F_i(\omega) = e^{j(\theta_i - \omega t_i)} \frac{b_i \tau_i}{\omega^2} \left(\operatorname{sinc} \frac{\omega \tau_i}{2} - \cos \frac{\omega \tau_i}{2} \right), \tag{4.17}$$

where $\operatorname{sinc} x = \dfrac{\sin x}{x}$. $F_i(\omega)$ is distributed over the whole frequency band from $\omega = -\infty$ to ∞. Fig. 4.1 shows an example of $F_i(\omega)$, which is the Fourier transform of a clipping pulse we arbitrarily selected from the simulation. The solid curve represents $|F_i(\omega)|$, and the dashed line illustrates the OFDM frequency band. We observe that $F_i(\omega)$ contributes a large portion of the out-of-band radiation. The in-band clipping noise is only a small portion of $F_i(\omega)$.

For multiple pulses, the PSD of the clipping noise is

$$S_f(\omega) = \frac{1}{T} E\{|F(\omega)|^2\} = \frac{1}{T} E\left\{ \sum_{i=1}^{N_p} |F_i(\omega)|^2 \right\} + \frac{1}{T} E\left\{ \sum_{i=1}^{N_p} \sum_{\substack{k=1 \\ k \neq i}}^{N_p} F_i(\omega) F_k^*(\omega) \right\}.$$

Because N_p is a random variable, we cannot directly exchange the order of summation

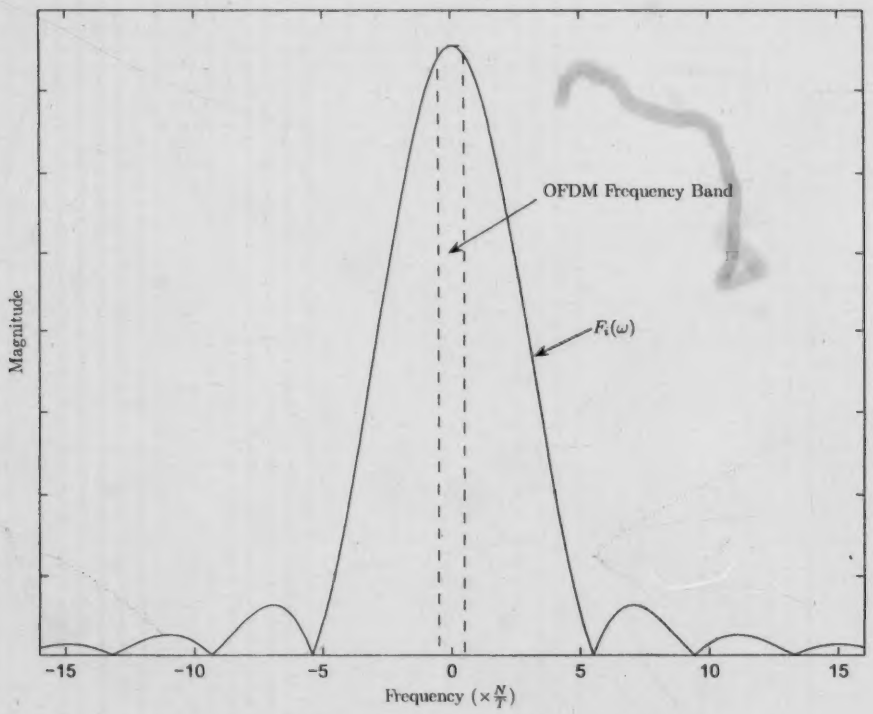


Figure 4.1: Frequency spectrum of $f_i(t)$.

and expectation. However, by the definition,

$$\begin{aligned}\frac{1}{T}E\left\{\sum_{i=1}^{N_p}|F_i(\omega)|^2\right\} &= \frac{1}{T}\lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n}\sum_{i=1}^{N_{c,l}}|F_{i,l}(\omega)|^2 \\ &= \frac{1}{T}\lim_{n\to\infty}\frac{\sum_{l=1}^{n}N_{c,l}}{n}\frac{\sum_{l=1}^{n}\sum_{i=1}^{N_{c,l}}|F_{i,l}(\omega)|^2}{\sum_{l=1}^{n}N_{c,l}} \\ &= \frac{\bar{N}_p}{T}E\{|F_i(\omega)|^2\} = \lambda_A E\{|F_i(\omega)|^2\},\end{aligned}$$

where the subscript l represents the l-th trial. Therefore,

$$S_f(\omega) = \lambda_A E\{|F_i(\omega)|^2\} + \frac{1}{T}E\left\{\binom{N_p}{2}\right\}E\{F_i(\omega)F_k^*(\omega)\},$$

where $i \neq k$. Note that

$$F_i(\omega)F_k^*(\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_i(\hat{t})f_k^*(\tilde{t})e^{-j\omega(\hat{t}-\tilde{t})}d\hat{t}d\tilde{t},$$

$E\{F_i(\omega)F_k^*(\omega)\}$ is determined by

$$E\{f_i(\hat{t})f_k^*(\tilde{t})\} = E\{(r(t_i+\Delta t_i)-A)e^{j\theta(t_i+\Delta t_i)}(r(t_k+\Delta t_k)-A)e^{-j\theta(t_k+\Delta t_k)}\}.$$

However, $r(t)$ and $\theta(t)$ are independent. Thus, we have

$$E\{f_i(\hat{t})f_k^*(\tilde{t})\} = E\{(r(t_i+\Delta t_i)-A)(r(t_k+\Delta t_k)-A)\}E\{e^{j\theta(t_i+\Delta t_i)}e^{j\theta(t_k+\Delta t_k)}\}.$$

We show in Appendix 4.7.A that, when $(t_i+\Delta t_i)$ and $(t_k+\Delta t_k)$ belong to different clipping pulses (as is true in our case), $x(t_i+\Delta t_i)$ and $x(t_k+\Delta t_k)$ are approximately independent, and thus $\theta(t_i+\Delta t_i)$ and $\theta(t_k+\Delta t_k)$ are uncorrelated. Then, $E\{F_i(\omega)F_k^*(\omega)\}=0$ and

$$S_f(\omega) = \lambda_A E\{|F_i(\omega)|^2\}. \tag{4.18}$$

The out-of-band radiation will be eliminated by filtering. Therefore, we are interested in the in-band clipping noise. When A is large, generally $\frac{\omega\tau_i}{2}$ is small for $|\omega| \le \frac{\pi N}{T}$. Thus, we may approximate $F_i(\omega)$ as

$$F_i(\omega) \approx e^{j(\theta_i-\omega t_i)}\frac{b_i\tau_i^3}{12} \tag{4.19}$$

by using [17] $\operatorname{sinc} x - \cos x \approx \frac{x^2}{3}$. Because $F_i(\omega)$ does not depend on the frequency ω, $S_f(\omega)$ is (approximately) constant over the OFDM band.

We may write b_i and τ_i as a function of $x_R(t_i)$, $x_I(t_i)$, $\dot{x}_I(t_i)$, $\ddot{x}_R(t_i)$, and $\ddot{x}_I(t_i)$. The joint pdf of these random variables can be easily found. However, a closed-form expression of $E\{b_i^2\tau_i^6\}$ cannot be obtained.

4.2.3 Clipping Noise Power Spectral Density

In this subsection, we calculate the in-band clipping noise PSD by using a result in [191]. Define $y(t) = \frac{r^2(t)}{\sigma^2}$, $\lambda = \frac{\dot{\sigma}^2}{\sigma^2}$ and $u = \frac{A^2}{\sigma^2}$. [191, Theorem 2.2] shows that, if $y(t)$ up-crosses the level u at $t=0$, with probability 1, $y(t)$ around $t=0$ can be written as

$$y(t) = -\lambda u t^2 + 2z\sqrt{\lambda u}t + u, \quad \text{when } u \to \infty, \tag{4.20}$$

where z is a Rayleigh random variable with the pdf

$$p(z) = ze^{-z^2/2}, \quad z > 0.$$

Also, the time duration τ between this up-crossing and the successive down-crossing is [191, Theorem 3.1]

$$\tau = \frac{2z}{\sqrt{\lambda u}}$$

with probability 1 when $u \to \infty$. Because, most probably, τ is very small for large A/σ in practical OFDM systems, we may use (4.20) to approximate the whole clipping pulse. Expanding $r(t) = \sigma\sqrt{y(t)}$ by using its Taylor's series at $t = 0$, we have

$$\begin{aligned} r(t) &\approx \sigma\sqrt{u} + \sqrt{\lambda}\sigma z t - \frac{1}{2}\lambda\sigma\left(\sqrt{u} + \frac{z^2}{\sqrt{u}}\right)t^2 \\ &\approx \sigma\sqrt{u} + \sqrt{\lambda}\sigma z t - \frac{1}{2}\lambda\sigma\sqrt{u}t^2 \\ &= -\frac{\dot{\sigma}^2 A}{2\sigma^2}\left(t - \frac{\tau}{2}\right)^2 + \frac{A\dot{\sigma}^2\tau^2}{8\sigma^2} + A, \quad 0 \le t \le \tau. \end{aligned} \tag{4.21}$$

The second step is obtained because $\dfrac{z^2}{\sqrt{u}} \ll \sqrt{u}$ when $u \to \infty$.

Remark 4.2. In the proof of [191, Theorem 2.2], $y(t/\sqrt{u})$ is first approximated by approximating $R(t)$ and its first derivative $\dot{R}(t)$, where $R(t)$ is the correlation function of $x_R(t)$ or $x_I(t)$, as polynomials of t with orders no larger than 2. Then, (4.20) is obtained for $u \to \infty$ by letting all terms that contain u^{-v}, where $v > 0$, be zero. In this paper, it is easy to check that the same approximation of $r(t)$ as in (4.21) is also obtained if $y(t/\sqrt{u})$ is approximated by approximating $R(t)$ and $\dot{R}(t)$ with polynomials of orders higher than 2.

Now, we can approximate a clipping pulse $f_k(t)$ that occurs in $t_k \le t \le t_k + \tau_k$ as

$$\begin{aligned} f_k(t) &= |r(t_k + \Delta t_k) - A|e^{j(\theta(t_k+\Delta t_k)} \\ &\approx \left(-\frac{\dot{\sigma}^2 A}{2\sigma^2}\left(\Delta t_k - \frac{\tau_k}{2}\right)^2 + \frac{A\dot{\sigma}^2\tau_k^2}{8\sigma^2}\right)e^{j(\theta_k+\eta_k\Delta t_k)}, \quad 0 \le \Delta t_k \le \tau, \end{aligned}$$

where $\theta_k = \theta(t_k)$, and

$$\begin{aligned} \eta_k &= \frac{\dot{x}_I(t_k)r(t_k) - x_I(t_k)\dot{r}(t_k)}{r(t_k)|x_R(t_k)|} \\ &= \frac{\dot{x}_I(t_k)A - x_I(t_k)\dot{r}(t_k)}{A|x_R(t_k)|}. \end{aligned}$$

Appendix 4.7.E shows that, when A is large, η_k has the same distribution as the γ_i used in the previous subsection. Thus, $\eta_k\Delta t_k$ is most probably small and can be ignored. Following the same procedure of the previous subsection, we have

$$\begin{aligned} F_k(\omega) &= e^{j(\theta_k - \omega(t_k+\tau_k/2))}\frac{\dot{\sigma}^2 A\tau_i}{\sigma^2(\omega-\eta_k)^2}\left(\operatorname{sinc}\frac{(\omega-\eta_k)\tau_k}{2} - \cos\frac{(\omega-\eta_k)\tau_i}{2}\right) \\ &\approx \frac{A\dot{\sigma}^2\tau_k^3}{12\sigma^2}e^{j(\theta_k - \omega(t_k+\tau_k/2))}. \end{aligned}$$

Then, the in-band clipping noise PSD is

$$\begin{aligned} S_f(\omega) &= \lambda_A E\{|F_k(\omega)|^2\} \\ &\approx \frac{A^3\dot{\sigma}^5}{144\sqrt{2\pi}\sigma^6} e^{-A^2/2\sigma^2} E\{\tau_k^6\} \\ &= \frac{32\sqrt{2}\sigma^6}{3\sqrt{\pi}\dot{\sigma}A^3} e^{-A^2/2\sigma^2} \\ &= \frac{32\sqrt{2}\sigma^2}{\pi\sqrt{3\pi}(A/\sigma)^3 W} e^{-A^2/2\sigma^2} \\ &= \frac{16\sqrt{2}}{\pi\sqrt{3\pi}(A/\sigma)^3} e^{-A^2/2\sigma^2} S_x, \end{aligned} \tag{4.22}$$

where $S_x = \dfrac{2\sigma^2}{W}$ is the PSD of the OFDM signal $x(t)$. For example, when $A/\sqrt{2}\sigma =$ 6 dB, the PSD of in-band clipping noise is -27 dB lower than that of the input OFDM signal.

4.2.4 Filtered Clipping Noise

The in-band clipping noise falls on both the reserved tones and the data tones. While the former must be kept for PAR reduction, the latter, as well as the out-of-band radiation, has to be filtered such that the clipping noise will not interfere with the data symbols or the communications on the neighboring frequency bands. In this subsection, we first consider reserving N_r consecutive tones around the center frequency; i.e., $\mathcal{R} = \{-\frac{N_r}{2}, -\frac{N_r}{2}+1, ..., \frac{N_r}{2}-1\}$. Other distributions of the reserved tones will be discussed later. To filter the clipping noise, we use an ideal lowpass filter with the passband $[-\omega_c, \omega_c]$ where

$$\omega_c = 2\pi f_c = 2\pi \frac{N_r}{2T} = \pi RW.$$

The filtered clipping noise is then given by

$$\hat{f}(t) = \frac{1}{2\pi} \int_{-\omega_c}^{\omega_c} \sum_{i=1}^{N_p} F_i(\omega) e^{j\omega t} d\omega. \tag{4.23}$$

By substituting (4.19) into (4.23), the filtered clipping noise may be expressed as

$$\hat{f}(t) = \hat{f}(t_i + \Delta t_i) = \sum_{i=1}^{N_p} \hat{f}_i(t) = \sum_{i=1}^{N_p} e^{j\theta_i} \frac{b_i \tau_i^3 f_c}{6} \operatorname{sinc} 2\pi f_c \Delta t_i. \tag{4.24}$$

Without loss of generality, we consider the clipping pulse $f_i(t)$ and assume it occurs at $t_i = 0$ and has the phase $\theta_i = 0$. Its filtered version is

$$\hat{f}_i(t) = \frac{b_i \tau_i^3 f_c}{6} \text{sinc}\, 2\pi f_c t. \tag{4.25}$$

Fig. 4.2 shows $f_i(t)$ and $\hat{f}_i(t)$. Several observations can be made by comparing $f_i(t)$ with $\hat{f}_i(t)$:

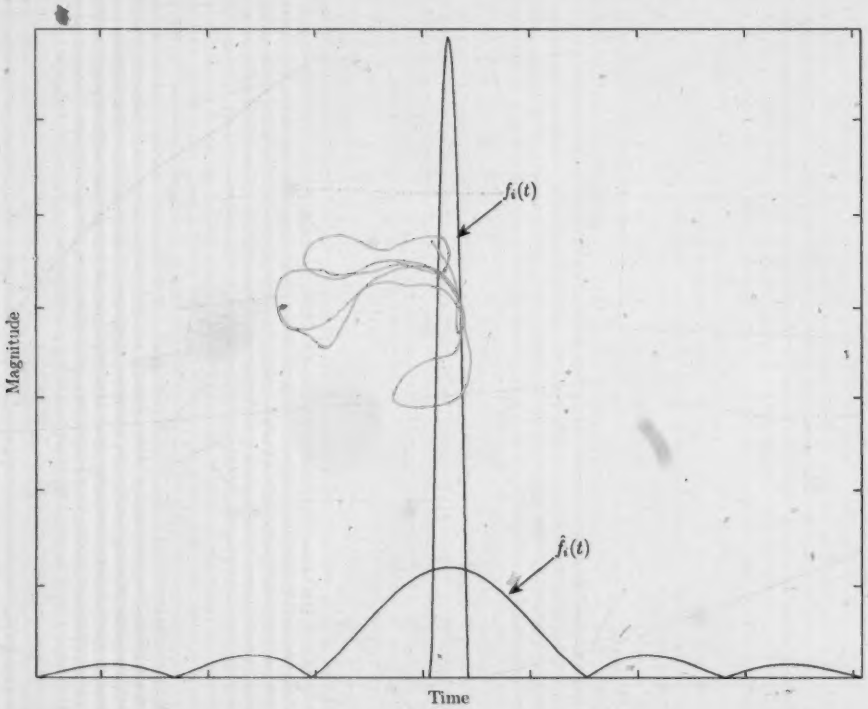


Figure 4.2: The clipping pulse $f_i(t)$ and its filtered version $\hat{f}_i(t)$.

1. $f_i(t)$ and $\hat{f}_i(t)$ reach their peaks at the same time instant $t = 0$.
2. $f_i(t)$ and $\hat{f}_i(t)$ have the same direction within the pulse duration of $f_i(t)$.
3. The mainlobe duration of $\hat{f}_i(t)$ is much wider than that of $f_i(t)$. The mainlobe duration of $\hat{f}_i(t)$ can be calculated as

$$\hat{\tau} = \frac{2T}{N_r}.$$

By using (4.11), the ratio of the average clipping pulse duration $\bar{\tau}$ over $\hat{\tau}$ is

$$\frac{\bar{\tau}}{\hat{\tau}} = \sqrt{\frac{3}{2\pi}} \frac{N_r}{N(\frac{A}{\sigma})} \ll 1$$

when A/σ is large.

4. The sidelobe peaks of $\left|\hat{f}_i(t)\right|$ decay with the rate of $1/t$. Specifically, the sidelobe peaks of $\left|\hat{f}_i(t)\right|$ are

$$\left|\hat{f}_i(T_k)\right| \approx \frac{b_i \tau_i^3 f_c}{3(2k+1)\pi} = \frac{2}{(2k+1)\pi}\left|\hat{f}_i(t)\right|_{\max}, \quad k = 1, 2, 3, ..., \tag{4.26}$$

with T_k representing the sidelobe peak occurrence time in accordance with

$$T_k \approx \frac{(2k+1)T}{2N_r} = \frac{2k+1}{2RW}. \tag{4.27}$$

For example, the peak of the first sidelobe is only 21.22% of that of the mainlobe.

5. The maximum of $\hat{f}_i(t)$ is much less than that of $f_i(t)$. In fact,

$$\left|\hat{f}_i(t)\right|_{\max} = \alpha\tau_i \left|f_i(t)\right|_{\max},$$

where α is defined as

$$\alpha = \frac{4}{3} f_c = \frac{2RW}{3}, \tag{4.28}$$

and the expectation of $\alpha\tau_i$ is

$$E\{\alpha\tau_i\} = \alpha\bar{\tau} = R\frac{2\sqrt{2}}{\sqrt{3\pi}}\frac{\sigma}{A} \ll 1$$

when A/σ is large. For example, $\alpha\bar{\tau} \approx 1.63\%$ when $A = 6\,\text{dB} \approx 1.9953$, $\sigma^2 = \frac{1}{2}$, and $R = 5\%$.

Point 5 explains why the peak regrows after filtering. Recall that the clipped signal is $\tilde{x}(t) = x(t) - f(t)$, and that after filtering it becomes $\hat{x}(t) = x(t) - \hat{f}(t)$. While $f(t)$ is chosen such that $|\tilde{x}(t_i)| = A$ at the peaks of $x(t)|_{t=t_i}$, the clipped and filtered signal $\hat{x}(t_i) > A$ because $\left|\hat{f}_i(t)\right|_{\max} < |f_i(t)|_{\max}$.

4.2.5 Iterative Clipping and Filtering: Single Clipping Pulse

To suppress the peak regrowth, clipping and filtering may be repeated until a suitable criterion is met. We first temporarily assume that the clipping noise at the first iteration $f^{(1)}(t)$ consists of only one dominant clipping pulse $f_i^{(1)}(t)$ (with pulse duration

$\tau_i^{(1)}$) that is much larger than other clipping pulses; i.e.,

$$f^{(1)}(t) = \sum_k f_k^{(1)}(t),$$
$$\left|f_i^{(1)}(t)\right| \gg \left|f_k^{(1)}(t)\right|, \quad \text{for all } k \neq i.$$

In this case, the clipped and filtered OFDM signal after the first iteration $\hat{x}^{(1)}(t) \approx x(t) - \hat{f}_i^{(1)}(t)$, where $\hat{f}_i^{(1)}(t)$ is the filtered version of $f_i^{(1)}(t)$, can be divided into three parts:

1) $|t| \leq \tau_i^{(1)}/2$.

Within this range, $\hat{f}_i^{(1)}(t)$, $f_i^{(1)}(t)$ and $x(t)$ have the same phase and $\left|\hat{f}_i^{(1)}(t)\right| < \left|f_i^{(1)}(t)\right|$. Therefore,

$$\left|\hat{x}^{(1)}(t)\right| \approx \left|x(t) - \hat{f}_i^{(1)}(t)\right| > \left|x(t) - f_i^{(1)}(t)\right| = A.$$

In other words, after passing $\hat{x}^{(1)}(t)$ through the SL, a clipping pulse, denoted as $f_i^{(2)}(t)$, occurs in the second clipping iteration at the same position as $f_i^{(1)}(t)$. By applying Taylor's series expansion to (4.25), and because $\frac{\omega_c \tau_i^{(1)}}{2} \ll 1$, we can approximate the filtered clipping pulse $\hat{f}_i^{(1)}(t)$ as a constant

$$\hat{f}_i^{(1)}(t) \approx \frac{b_i(\tau_i^{(1)})^3 f_c}{6} = \left|\hat{f}_i^{(1)}(t)\right|_{\max}, \quad |t| \leq \frac{\tau_i^{(1)}}{2}.$$

Then, the clipping pulse at the second iteration $f_i^{(2)}(t)$ can be written as

$$f_i^{(2)}(t) = f_i^{(1)}(t) - \hat{f}_i^{(1)}(t) \approx -\frac{1}{2}b_i t^2 + \frac{1}{8}b_i(\tau_i^{(1)})^2 - \frac{b_i(\tau_i^{(1)})^3 f_c}{6}, \tag{4.29}$$

which is also a parabolic arc with reduced magnitude. By solving $f_i^{(2)}(t) = 0$, the time duration of $f_i^{(2)}(t)$ can be found as

$$\tau_i^{(2)} = \tau_i^{(1)}\sqrt{1 - \frac{4}{3}\tau_i^{(1)} f_c} = \tau_i^{(1)}\sqrt{1 - \alpha^{(1)}\tau_i^{(1)}},$$

where $\alpha^{(1)}$ is the α defined in (4.28).

2) $\frac{\tau_i^{(1)}}{2} < |t| < T_2$, where T_2 is given in (4.27).

In this range, $|x(t)| < A$ because only one clipping pulse exists. However, depending on the phase of $\hat{f}_i^{(1)}(t)$, $\left|\hat{x}^{(1)}(t)\right|$ may be greater than A. In other words, new clipping pulses may be generated in the second clipping iteration. However, because

$\left|\hat{f}_i^{(1)}(t)\right|_{\max} \ll \left|f_i^{(1)}(t)\right|_{\max}$, these new clipping pulses are very small compared to the clipping pulse $f_i^{(2)}(t)$, and their effects can be ignored.

3) $|t| > T_2$.

Because the peaks of $\hat{f}_i^{(1)}(t)$ decay with the rate of $1/t$, we can see that, in this range, $\hat{x}^{(1)}(t) \approx x(t)$. Therefore, no clipping pulses exist at $|t| > T_2$ in the second clipping iteration.

The successive clipping and filtering iterations repeat this procedure. Therefore, we conclude:

For the case of only one dominant clipping pulse, in the l-th ($l = 2, 3, ...$) clipping and filtering iteration, $f_i^{(l-1)}(t)$ shrinks to $f_i^{(l)}(t)$, and some new pulses possibly appear. Here, $f_i^{(l-1)}(t)$ and $f_i^{(l)}(t)$ are the dominant clipping pulses at the $(l-1)$-th and l-th iterations, respectively. Until $f_i^{(l)}(t)$ is comparable to the new pulses, the latter can be omitted, and the former can be written as

$$f_i^{(l)}(t) = -\frac{1}{2}b_i t^2 + \frac{1}{8}b_i(\tau_i^{(l)})^2, \quad i = 1, 2, 3, ..., \tag{4.30}$$

where

$$f_i^{(1)}(t) = f_i(t),$$
$$\tau_i^{(1)} = \tau_i,$$

and

$$\tau_i^{(l)} = \tau_i^{(l-1)}\sqrt{1 - \alpha\tau_i^{(l-1)}}, \quad i = 2, 3, 4, ..., \tag{4.31}$$

and α is defined in (4.28). Moreover, the filtered clipping pulse in the l-th iteration is

$$\hat{f}_i^{(l)}(t) = \frac{b_i(\tau_i^{(l)})^3 f_c}{6}\operatorname{sinc} 2\pi f_c t. \tag{4.32}$$

Thus, the filtered clipping noise generated in the l-th iteration is proportional to that generated in the first iteration. Define β as

$$\beta \triangleq \frac{\text{total filtered clipping noise after } K \text{ iterations}}{\text{filtered clipping noise generated in the first iteration}}.$$

If only one dominant clipping pulse exists,

$$\beta = \frac{\sum_{l=1}^{K}\hat{f}_i^{(l)}(t)}{\hat{f}_i^{(1)}(t)} = \frac{\sum_{l=1}^{K-1}(\tau_i^{(l)})^3}{(\tau_i^{(1)})^3}. \tag{4.33}$$

Finding $\bar{\beta}$, the mean of β, is difficult. However, an estimation of $\bar{\beta}$ can be obtained when K is not large. When A is large, $\alpha\tau_i^{(l)} \ll 1$. Then, $\sqrt{1-\alpha\tau_i^{(l)}}$ can be treated as a constant. $\bar{\beta}$ can then be estimated by replacing $\tau_i^{(1)}$ with its mean $\bar{\tau}$. Thus,

$$\bar{\beta} \approx \frac{1-(1-\alpha\bar{\tau})^{3K/2}}{1-(1-\alpha\bar{\tau})^{3/2}}. \tag{4.34}$$

We will use $\bar{\beta}$ in the constant-scaling algorithm proposed in the next section.

Remark 4.3. Because $\pi\tau_i W$ is most probably very small when A is large, it is easy to show that (4.33) and (4.34) also apply to other reserved tone sets.

4.2.6 Iterative Clipping and Filtering: Multiple Pulses

In OFDM systems, the (unfiltered) clipping noise is usually a series of parabolic pulses. In each clipping and filtering iteration, the filtered clipping noise at $t_i - \tau_i^{(l)}/2 \le t \le t + \tau_i^{(l)}/2$ is the mainlobe of $\hat{f}_i^{(l)}(t)$ plus the mainlobes (if they are close to $t = t_i$) or the sidelobes (if they are far from $t = t_i$) of all the other filtered pulses at $t_i - \tau_i^{(l)}/2 \le t \le t + \tau_i^{(l)}/2$.

If a clipping pulse contributes an impact stronger than its k-th sidelobe, where k is properly chosen such that the k-th sidelobe is relatively large and cannot be omitted, the clipping pulse must occur within the time interval $t_i - T_k \le t \le t_i + T_k$. Eq. (4.57) in Appendix 4.7.A gives the probability that two or more clipping pulses occur within a time interval. Substituting (4.27) into (4.57), we can see that such a probability is independent of N. In other words, with a fixed probability, the number of clipping pulses that occur within the time interval $t_i - T_k \le t \le t_i + T_k$ is independent of N. However, while $N \to \infty$, the average number of clipping pulses also goes to ∞ (see (4.9)). Therefore, the effect of most other filtered pulses at $t_i - \tau_i^{(l)}/2 \le t \le t + \tau_i^{(l)}/2$ can be omitted. We need to consider only the pulses close to the mainlobe of $\hat{f}_i^{(l)}(t)$.

After the l-th iteration, the peak-reduced OFDM signal at $t = t_i$ becomes

$$\hat{x}^{(l+1)}(t_i) = (A + |f_i^{(l)}(t_i)| - |\hat{f}_i^{(l)}(t_i)|)e^{j\theta^{(l)}(t_i)} - \sum_{m \ne i} \hat{f}_m^{(l)}(t_i),$$

where $\theta^{(l)}(t_i)$ is the phase of $\hat{x}^{(l)}(t_i)$. When A is very large, say $A \ge 9\,\text{dB}$, all $f_m^{(l)}(t)$ are far apart from $f_i^{(l)}(t)$, and $|\hat{f}_m^{(l)}(t_i)| \approx 0$ for all $m \ne i$. In this case, our conclusions

in the previous section hold. It is easy to show that, with $l \to \infty$, the clipping noise at the l-th iteration $f^{(l)}(t) \to 0$, and the peak of $x(t)$ at $t = t_i$ is reduced to A.

On the other hand, for moderate A, some of $\hat{f}_m^{(l)}(t_i)$ may be relatively large and cannot be omitted. In this case, we decompose $\hat{f}_m^{(l)}(t_i)$ as

$$\hat{f}_m^{(l)}(t_i) = (\hat{f}_{m,I}^{(l)}(t_i) + j\hat{f}_{m,Q}^{(l)}(t_i))e^{j\theta^{(l)}(t_i)},$$

where $\hat{f}_{m,I}^{(l)}(t_i)$ and $\hat{f}_{m,Q}^{(l)}(t_i)$ are the inphase and quadrature components along the direction of $\hat{x}^{(l)}(t_i)$, respectively. By noting that, most probably, A is much larger than any clipping pulse, $\hat{f}_{m,Q}^{(l)}(t_i)$ can be omitted when calculating $\hat{x}^{(l+1)}(t_i)$. Then, in the $(l+1)$-th iteration, the clipping pulse at $t = t_i$ is

$$f_i^{(l+1)}(t_i) \approx \left(|f_i^{(l)}(t_i)| - |\hat{f}_i^{(l)}(t_i)| - \sum_{m \neq i} \hat{f}_{m,I}^{(l)}(t_i) \right) e^{j\theta^{(l)}(t_i)}.$$

Depending on the sign of $\sum_{m \neq i} \hat{f}_{m,I}^{(l)}(t_i)$, the peak reduction may be strengthened or weakened. Moreover, $|f_i^{(l+1)}(t_i)| > |f_i^{(l)}(t_i)|$; i.e., the peak is *increased*, when $\sum_{m \neq i} \hat{f}_{m,I}^{(l)}(t_i) < -|\hat{f}_i^{(l)}(t_i)|$.

Remark 4.4. In the multiple clipping pulses case, the validity of (4.33) and (4.34) depends on A. That is, (4.33) and (4.34) are valid when A is large. Otherwise, the estimation error of these two equations is relatively large. However, the above analysis is still valid, and the total filtered clipping noise is still proportional to that generated in the first iteration until A is so small that the width of the mainlobe of $F_i(\omega)$ is comparable to or smaller than W.

Remark 4.5. A wider mainlobe of a filtered clipping pulse implies that its magnitude will be interfered with by more neighboring clipping pulses. This interference, in turn, implies a worse PAR-reduction performance. Therefore, the filtered clipping pulse should be made close to an impulse function; i.e., both the width of the mainlobe and the magnitudes of the sidelobes must be minimized.

We will now check at what level of A (4.33) and (4.34) are valid. Such a level of A, denoted as A_{thres}, depends on the choice of the reserved-tone set. [18] proves that a close-to-optimum reserved-tone set can be found from a small number of randomly selected reserved-tone sets. This proof implies that consecutive reserved tones usually

lead to non-optimal solutions. Therefore, we may use the consecutive reserved tones to find A_{thres}. In other words, if (4.33) and (4.34) are valid for consecutive reserved tones when $A \geq A_{\text{thres}}$, they are also valid for most other reserved-tone sets when $A \geq A_{\text{thres}}$.

Eq. (4.27) indicates that the width of the mainlobe and sidelobes of the filtered clipping pulse $\hat{f}_i(t)$ is determined only by the tone reservation ratio R and the OFDM bandwidth W. Therefore, we may use a T_k as a reference such that the tail of $\hat{f}_i(t)$ beyond T_k is small and can be ignored. From (4.26), the peak of the 4th sidelobe is only 7% of that of the mainlobe. Therefore, if any pair of pulses are apart by at least T_4 seconds, (4.33) and (4.34) are valid. By substituting T_4 into (4.59), the probability that more than one clipping pulse occur with a time duration of T_4, given that a clipping pulse has already occurred in this time interval, is

$$\Pr(1) = 1 - e^{-T_4 \lambda_A}.$$

Fig. 4.3 illustrates the relationship of $\Pr(1)$ and A for $R = 0.05$, 0.1 and 0.2, where $\sigma = 1/\sqrt{2}$ and T_4 are used. In this figure, we also include $R = 1$ as a reference. We refer to $R = 1$ as the iterative clipping and filtering technique [12] where no tone is reserved and clipping noise is distributed over the whole OFDM band.

We see that, for $R = 0.05$, 0.1 and 0.2, (4.33) and (4.34) can be used with a small approximation error when $A \geq 9\,\text{dB}$, $8.5\,\text{dB}$ and $8\,\text{dB}$, respectively. However, if no tone is reserved and the clipping noise is distributed over the whole OFDM band ($R = 1$), the two equations are valid when $A \geq 6.5\,\text{dB}$.

Remark 4.6. Here, we choose a strict criterion to find A_{thres}. A less strict criterion can be obtained by selecting a larger sidelobe (e.g., T_1) as the reference, by using a $\Pr(m)$ where $m > 1$, and/or allowing a larger $\Pr(m)$ in determining A_{thres}.

4.2.7 Effect of Reserved-Tone Position on PAR Reduction

Let us consider using an ideal bandpass filter with passbands on only the reserved tones $\mathcal{R}$ to filter the i-th clipping pulse $f_i(t)$.[2] We assume that $f_i(t)$ occurs at $t_i = 0$

[2]Although $\mathcal{R}$ is a discrete set, here, for simplicity, we slightly misuse $\mathcal{R}$ to represent the reserved tones in the continuous frequency domain. In this case, each item $i \in \mathcal{R}$ represents a frequency band with width $\frac{1}{T}$ and the central frequency $\frac{i}{T}$.

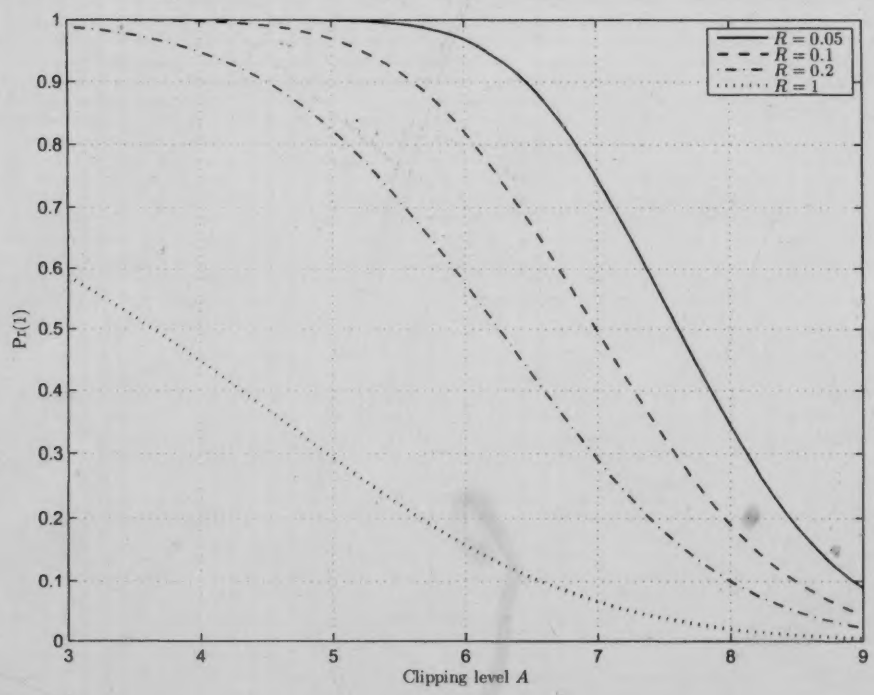


Figure 4.3: The probability that more than one clipping pulses occur with a time duration of T_4, given a clipping pulse has already occurred in this time interval, where $\sigma = 1/\sqrt{2}$.

and has a phase $\theta_i = 0$. The filtered clipping pulse is

$$\hat{f}_i(t) = \frac{1}{2\pi} \int_{\mathcal{R}} \frac{b_i(\tau_i)^3}{12} e^{j\omega t} d\omega = \frac{b_i(\tau_i)^3}{12T} h(t), \tag{4.35}$$

where

$$h(t) = \text{sinc}\, \frac{\pi t}{T} \sum_{k \in \mathcal{R}} e^{j2\pi k \frac{t}{T}} \tag{4.36}$$

is the impulse response of the bandpass filter, and[3] $-T \leq t \leq T$ because the time duration of an OFDM symbol is T. When $\mathcal{R} = \{-\frac{N_r}{2}, -\frac{N_r}{2} + 1, ..., \frac{N_r}{2} - 1\}$, (4.35) reduces to (4.25).

To make $\hat{f}_i(t)$ close to an impulse function, we need to consider only the magnitude response $|h(t)|$. Some conclusions can be obtained from (4.36):

1) The mainlobe of $h(t)$ is of the maximum width when $N_r = 1$, implying the worst PAR-reduction capability.

[3]The actual range of t is $-t_k \leq t \leq T - t_k$. Considering $0 \leq t_k \leq T$, we have $-T \leq t \leq T$.

2) For consecutive reserved tones, the mainlobe of $h(t)$ is of the minimum width when $N_r = N$, implying the best PAR-reduction capability.

If we choose another set of reserved tones $\mathcal{R}'$, which is a shift of $\mathcal{R}$; i.e., $\mathcal{R}' = \mathcal{R}+n_s$, where $\mathcal{R}$ is of any kind, and n_s is an integer, then the impulse response of $\mathcal{R}'$ is

$$\begin{aligned} |h'(t)| &= \left| \text{sinc}\, \frac{\pi t}{T} \sum_{k\in\mathcal{R}'} e^{j2\pi k \frac{t}{T}} \right| = \left| e^{j2\pi n_s \frac{t}{T}} \text{sinc}\, \frac{\pi t}{T} \sum_{k\in\mathcal{R}} e^{j2\pi k \frac{t}{T}} \right| \\ &= |h(t)|. \end{aligned} \tag{4.37}$$

That is, $h'(t)$ has the same magnitude response as $h(t)$.

3) Therefore, shifting $\mathcal{R}$ cannot change the PAR-reduction capability.

In (4.36), the term $\text{sinc}\, \frac{\pi t}{T}$ renders the envelope of $h(t)$, and the width of the mainlobe of $h(t)$ is determined by

$$g(t) = \sum_{k\in\mathcal{R}} e^{j2\pi k \frac{t}{T}}.$$

Let $\mathbf{b} = [b_{-\frac{N}{2}}, b_{-\frac{N}{2}+1}, ..., b_{\frac{N}{2}-1}]$ be the indicator of reserved tones; i.e.,

$$b_k = \begin{cases} 1, & k \in \mathcal{R}, \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$g(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} b_k e^{j2\pi k \frac{t}{T}}.$$

Because the phase of $h(t)$ is irrelevant to our consideration, we can focus on $|g(t)|^2$, but

$$|g(t)|^2 = \sum_{n=-N+1}^{N-1} \Psi(n) e^{j2\pi nt/T},$$

where

$$\Psi(n) = \begin{cases} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1-n} b_k b_{k+n}, & n \geq 0, \\ \sum_{k=-\frac{N}{2}-n}^{\frac{N}{2}-1} b_k b_{k+n}, & n < 0 \end{cases}$$

is the aperiodic autocorrelation function of $\mathbf{b}$. The width of $\Psi(n)$ is inversely proportional to that of $|g(t)|^2$.

4) Then, for fixed N_r, the consecutive reserved-tone positioning leads to a small PAR reduction because the consecutive reserved-tone positioning has the narrowest

width of $\Psi(n)$ compared to that of other positioning schemes. In other words, a randomly chosen reserved-tone set usually leads to a better PAR-reduction performance than the consecutive reserved-tone set [18].

Minimizing only the mainlobe width of $h(t)$ does not maximize the PAR-reduction performance. In fact, the minimum width of $h(t)$ can be obtained by maximizing the width of $\Psi(n)$; i.e., putting reserved tones at the both ends of the OFDM frequency band. In other words,

$$\mathcal{R} = \{-\frac{N}{2}, -\frac{N}{2}+1, ..., -\frac{N}{2}+\frac{N_r}{2}-1\} \cup \{\frac{N}{2}-\frac{N_r}{2}, \frac{N}{2}-\frac{N_r}{2}+1, ...\frac{N}{2}-1\}.^4 \quad (4.38)$$

Fig. 4.4 (a) illustrates this positioning scheme. The corresponding $h(t)$ is shown in Fig. 4.4 (b) as the solid curve. As a reference, a randomly chosen positioning scheme is also shown in Fig. 4.4 (b) as the dashed curve. Although the minimum-width positioning scheme of (4.38) leads to the minimum width of $h(t)$, it also leads to significantly larger sidelobes than the randomly chosen positioning scheme. Therefore, the minimum-width positioning scheme cannot maximize PAR reduction performance.

To find the optimal positioning scheme, we may optimize $\mathcal{R}$ such that the maximum peak in $|t| > \frac{\bar{\tau}}{2}$ is minimized; i.e.,

$$\min_{\mathcal{R}} \max_{T \geq |t| > \frac{\bar{\tau}}{2}} |h(t)|, \quad (4.39)$$

which is equivalent to

$$\begin{aligned} &\min_{\mathbf{b}} \max_{T \geq |t| > \frac{\bar{\tau}}{2}} \left| g(t) \text{sinc} \frac{\pi t}{T} \right|, \\ &\text{subject to: } \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} b_k = N_r. \end{aligned} \quad (4.40)$$

$\text{sinc} \frac{\pi t}{T}$ is only a weighting factor, and $g(t)$ is periodic with a period of T. Moreover, $|g(t)| = |g(-t)|$ and $\left|\text{sinc} \frac{\pi t}{T}\right| = \left|\text{sinc}\left(-\frac{\pi t}{T}\right)\right|$ because b_i are integers. Then, (4.40) is

[4] (4.38) is valid only for the zero-inserting scheme. For the zero-padding scheme, which pads $JN - N$ zeros at the end of an OFDM block, the $\mathcal{R}$ in (4.38) will be $\mathcal{R} = \{0, 1, ..., \frac{N_r}{2}-1\} \cup \{N - \frac{N_r}{2}+1, N-\frac{N_r}{2}+2, ..., N\}$.

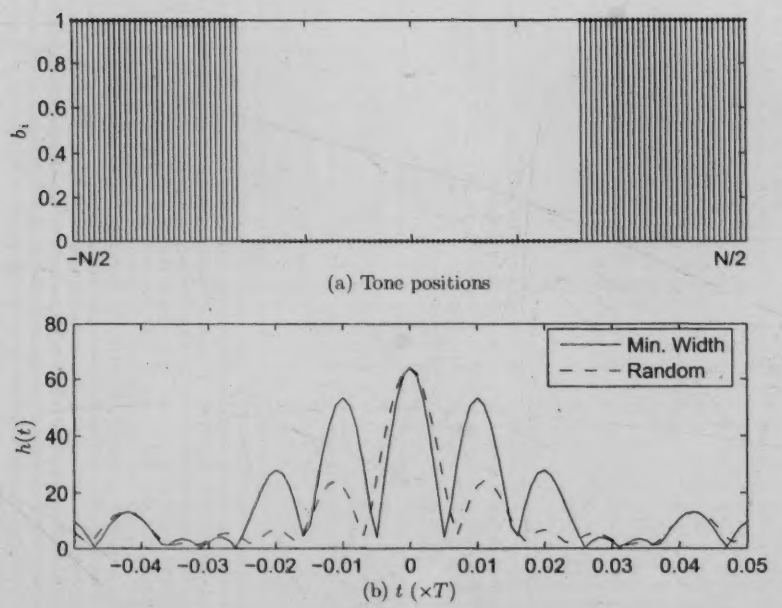


Figure 4.4: Tone positions and the corresponding impulse response $h(t)$.

equivalent to

$$\begin{aligned} &\min_{\mathbf{b}} \max_{\frac{T}{2} \geq |t| > \frac{\tau}{2}} \left| g(t) \text{sinc} \frac{\pi t}{T} \right|, \\ &\text{subject to: } \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} b_k = N_r. \end{aligned} \tag{4.41}$$

Sampling the objective function with the sampling frequency $\frac{JN}{T}$, where J is the oversampling factor, (4.41) becomes

$$\begin{aligned} &\min_{\mathbf{b}} \max_{n_r \leq n \leq \frac{JN}{2}-1} \left| g(n) \text{sinc} \frac{\pi n}{JN} \right|, \\ &\text{subject to: } \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} b_k = N_r, \end{aligned} \tag{4.42}$$

where

$$g(n) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} b_k e^{j2\pi \frac{kn}{JN}}, \quad n = 0, 1, ..., \frac{JN}{2} - 1,$$

and

$$n_\tau = \left\lceil J\sqrt{\frac{6}{\pi}}\frac{\sigma}{A} \right\rceil,$$

with $\lceil x \rceil$ representing the minimum integer that is greater than x. If the factor $\text{sinc}\,\frac{\pi n}{JN}$ is omitted, and we let $n_\tau = J$, (4.42) reduces to the optimization of $\mathcal{R}$ proposed in [18].

The search space of (4.42) is $\binom{N}{N_r}$, which is prohibitively large when N and N_r are large. Alternatively, we can use the random set optimization [18] to find a suboptimal solution $\mathcal{R}^*$, i.e., to select the best from M randomly generated reserved-tone sets $\mathcal{R}_1, ..., \mathcal{R}_M$.

4.3 New Tone-Reservation Algorithms

Based on our analysis of clipping noise, we propose two new tone-reservation algorithms for PAR reduction. The main idea is that because the clipping noise in each iteration is similar, the total filtered clipping noise can be approximated by scaling the filtered clipping noise generated in the first iteration.

4.3.1 Constant Scaling Tone Reservation Algorithm

If a relatively high PAR is tolerable, but fast execution time is a must, we propose a constant-scaling tone-reservation algorithm. This algorithm scales the filtered clipping noise by a constant factor $\bar{\beta}$, and subtracts the scaled clipping noise from the original OFDM symbol. This algorithm can be stated as follows:

Algorithm 4.1 (Constant-Scaling).

Initialization:

This stage needs to run only once.

1. Choose a relatively high clipping threshold A; randomly choose the reserved-tone set $\mathcal{R}$, or set it up by using random set optimization.

2. Choose a K, and calculate $\bar{\beta}$ by using (4.34).

Runtime:

1. Distribute $(N - N_r)$ input symbols to data tones $\mathcal{R}^c$, and calculate the corresponding time domain signal x_n by using (4.2). Oversampling may be required.

2. If PAR $> A$, go to step 3; otherwise, transmit x_n and terminate.

3. Clip x_n to the threshold A to find the clipping noise f_n by using the discrete-time version of (4.4).

4. Filter f_n subject to tone-reservation and other constraints, and obtain the peak-canceling signal c_n. The filtering can be done by using a pair of Discrete Fourier Transform (DFT)/Inverse Discrete Fourier Transform (IDFT) as follows:

   4-a. Convert f_n to the frequency domain by using a DFT to obtain $F_k = \text{DFT}\{\mathbf{f}\}$, where $\mathbf{f} = [f_0, ..., f_{JN-1}]$.

   4-b. Keep the first and last $\frac{N}{2}$ items of F_k to obtain the in-band distortion $\tilde{\mathbf{F}}$; i.e.,

   $$\tilde{\mathbf{F}} = [\tilde{F}_0, \tilde{F}_1, ..., \tilde{F}_{N-1}] = [F_0, F_1, ..., F_{\frac{N}{2}-1}, F_{JN-\frac{N}{2}}, F_{JN-\frac{N}{2}+1}, ..., F_{JN-1}].$$

   4-c. The filtered clipping noise is $\hat{\mathbf{F}} = [\hat{F}_0, \hat{F}_1, ..., \hat{F}_{N-1}]$, where

   $$\hat{F}_k = \begin{cases} \tilde{F}_k, & k \in \mathcal{R}, \\ 0, & \text{otherwise.} \end{cases} \tag{4.43}$$

   Then, by scaling $\hat{F}_k$, the peak reduction signal C_k becomes $C_k = -\beta \hat{F}_k$. Scaling the filtered clipping noise in the frequency domain involves less arithmetic operations than scaling it in the time domain.

   4-d. Convert C_k to time domain to obtain c_n by using an IDFT.

5. Calculate the PAR-reduced OFDM signal as $\hat{x}_n = x_n + c_n$, and transmit it.

Remark 4.7. Calculating the PAR and finding f_n require the calculation of $|x_n|$, which is costly if all $|x_n|$ are calculated. Here, we propose a method to reduce such a cost. We need to calculate only $|x_n|$ for those $|x_n| \geq A$. A necessary condition of $|x_n| \geq A$ is

$$\left(|x_{n,R}| \geq \frac{A}{\sqrt{2}} \text{ OR } |x_{n,I}| \geq \frac{A}{\sqrt{2}}\right) \text{ AND } (|x_{n,R}| + |x_{n,I}| \geq A), \tag{4.44}$$

where $x_{n,R}$ and $x_{n,I}$ are the real and imaginary parts of x_n, respectively. Thus, we need to calculate only $|x_n|$ for the samples satisfying (4.44). Later, we will show that the number of the samples satisfying (4.44) is small. This method can also be used in other PAR-reduction techniques such as SLM and PTS.

4.3.2 Adaptive-Scaling Tone-Reservation Algorithm

In this subsection, we propose an adaptive-scaling tone-reservation algorithm to obtain a large PAR reduction. Instead of using $\bar{\beta}$, this algorithm calculates β for each OFDM symbol.

A discrete-time domain description of the algorithm is given here. Filtering the clipping noise f_n to $\mathcal{R}$, we get the filtered clipping noise $\hat{f}_n = \text{IDFT}\{\hat{\mathbf{F}}\}$, where $\hat{\mathbf{F}}$ is obtained by using (4.43). The PAR-reduced signal $\hat{x}_n$ can be written as

$$\hat{x}_n = x_n - \beta\hat{f}_n = Ae^{j\theta_n} + f_n - \beta\hat{f}_n, \tag{4.45}$$

where θ_n is the phase of x_n. Our task is to minimize the out-of-range power P, i.e., the total power of those $|\hat{x}_n| > A$. The objective function is

$$\min_{\beta} P, \tag{4.46}$$

where

$$P = \sum_{|\hat{x}_n|>A} (|\hat{x}_n| - A)^2. \tag{4.47}$$

Eq. (4.47) can be rewritten as

$$P = \sum_{n\in\mathcal{S}_1} (|\hat{x}_n| - A)^2 - \sum_{\substack{n\in\mathcal{S}_t \\ |\hat{x}_n|\le A}} (|\hat{x}_n| - A)^2 + \sum_{n\in\mathcal{S}_2} (|\hat{x}_n| - A)^2, \tag{4.48}$$

where $\mathcal{S}_1 = \{n : |f_n| > 0\}$ is the index set of all clipping pulses, and $\mathcal{S}_2 = \{n : |f_n| = 0 \text{ and } |\hat{x}_n| > A\}$. Because clipping pulses are parabolic arcs, the power of any clipping pulse is a monotonic function of its peak amplitude. Minimizing (4.48) is equivalent to minimizing

$$\hat{P} = \underbrace{\sum_{n\in\mathcal{S}_p} (|\hat{x}_n| - A)^2}_{P_1} - \underbrace{\sum_{\substack{n\in\mathcal{S}_p \\ |\hat{x}_n|\le A}} (|\hat{x}_n| - A)^2}_{P_2} + \underbrace{\sum_{n\in\mathcal{S}_p^+} (|\hat{x}_n| - A)^2}_{P_3}, \tag{4.49}$$

where $\mathcal{S}_p = \{n : n \in \mathcal{S}_1, |x_n| > |x_{n-1}|$, and $|x_n| \geq |x_{n+1}|\}$ is the index set of the peaks of f_n; $\mathcal{S}_p^+ = \{n : n \in \mathcal{S}_2, |\hat{x}_n| > |\hat{x}_{n-1}|$, and $|\hat{x}_n| \geq |\hat{x}_{n+1}|\}$ is the index set of the peaks of newly generated pulses whose amplitudes are larger than A. That is, if $\hat{P}$ is minimized, P is also close-to-optimally minimized.

Eq. (4.49) implies that the optimal β, denoted as $\beta^{(\text{opt})}$, must both minimize the peaks of x_n and prevent any large newly-generated pulses. This implication, in turn, implies that β cannot be large. Thus, P_2 and P_3 are small and their difference can be omitted. Therefore,

$$\begin{aligned} \hat{P} \approx P_1 &= \sum_{n \in \mathcal{S}_p} |\hat{x}_n - Ae^{j\hat{\theta}_n}|^2 \\ &= \sum_{n \in \mathcal{S}_p} |f_n - \beta \hat{f}_n + A(e^{j\theta_n} - e^{j\hat{\theta}_n})|^2, \end{aligned} \tag{4.50}$$

where $\hat{\theta}_n$ is the phase of $\hat{x}_n$.

Because $\beta^{(\text{opt})}$ is not large, we can see that $|x_n| = |Ae^{j\theta_n} + f_n| \gg |\beta \hat{f}_n|$, i.e., $\beta \hat{f}_n$ could not significantly change the phase of x_n. Therefore, $\hat{\theta}_n \approx \theta_n$, and

$$\hat{P} \approx \sum_{n \in \mathcal{S}_p} |f_n - \beta \hat{f}_n|^2. \tag{4.51}$$

The optimal solution is

$$\beta^{(\text{opt})} = \frac{\mathfrak{Re}[\sum_{n \in \mathcal{S}_p} f_n \hat{f}_n^*]}{\sum_{n \in \mathcal{S}_p} |\hat{f}_n|^2}, \tag{4.52}$$

where $\mathfrak{Re}[x]$ represents the real part of x, and $(\cdot)^*$ represents the complex conjugate.

If β can be a complex number, the optimal solution is

$$\beta_c^{(\text{opt})} = \frac{\sum_{n \in \mathcal{S}_p} f_n \hat{f}_n^*}{\sum_{n \in \mathcal{S}_p} |\hat{f}_n|^2}.$$

However, $\beta^{(\text{opt})}$ is determined mainly by some dominant peaks. Most likely, if f_n is a dominant peak, $\hat{f}_n$ is also large, and the phase of $\hat{f}_n$ is close to that of f_n. Therefore, the imaginary part of $\beta_c^{(\text{opt})}$ is small and can be omitted.

To further reduce the execution time, some small samples of f_n can be excluded from (4.52). Doing so, however, may degrade the PAR-reduction performance.

Now, the adaptive-scaling tone-reservation algorithm can be summarized as follows:

Algorithm 4.2 (Adaptive-Scaling).

Initialization:

1. Set up A, $\mathcal{R}$, and the maximum number of iterations L.

Runtime:

For each length $(N - N_r)$ OFDM block,

1. Find x_n.
2. If PAR $> A$, go to step 3; otherwise, transmit x_n and terminate the algorithm.
3. Find the clipping noise f_n.
4. Find c_n as follows:
   - 4-a. Find the filtered clipping noise (in the frequency domain) $\hat{\mathbf{F}}$ as in Algorithm 4.1. Here, a DFT on $\mathbf{f}$ is needed.
   - 4-b. Convert $\hat{\mathbf{F}}$ to the time domain to obtain $\hat{f}_n = \text{IDFT}(\hat{\mathbf{F}})$.
   - 4-c. Find the peaks of f_n to obtain $\mathcal{S}_p$.
   - 4-d. Find $\beta^{(\text{opt})}$ by using (4.52).
   - 4-e. The peak reduction signal is $c_n = -\beta^{(\text{opt})} \hat{f}_n$.
5. Calculate the PAR-reduced OFDM signal as $\hat{x}_n = x_n + c_n$. If PAR $> A$, and the iteration number is less than L, go to Step 3. Otherwise, transmit $\hat{x}_n$, and terminate the algorithm.

4.3.3 Complexity Comparison

4.3.3.1 Complexity Analysis of the Proposed Algorithms

We now consider the complexity of our algorithms by calculating the number of real multiplications. Here, we count a complex multiplication as three real multiplications [192]. Only the runtime computational cost is considered, and the cost of the initialization stage can be omitted because it occurs only once. In the runtime stage of both algorithms, Steps 1 and 2 are not counted either because all OFDM systems

must execute Step 1, and all PAR-reduction techniques require at least one iteration of these two steps.

In Step 3, f_n can be calculated as $f_n = x_n(1-\frac{A}{|x_n|})$, where $n \in \mathcal{F} = \{n : |x_n| > A\}$, and N_f is the size of $\mathcal{F}$. The execution time of this step is determined by that of calculating $|x_n|$ and f_n.

By applying the condition (4.44) to exclude small samples, the execution time of calculating $|x_n|$ is small. The number of samples that satisfy (4.44) is

$$\bar{N}_c = JN\left(1 - (\mathrm{erf}\,(A/2\sigma))^2 - P1\right),$$

where

$$P1 = \int_{A/\sqrt{2}}^{A} \frac{2\sqrt{2}}{\sigma\sqrt{\pi}} \mathrm{erf}\left(\frac{A-x}{\sigma\sqrt{2}}\right) e^{-x^2/2\sigma^2} dx.$$

For example, $\bar{N}_c \approx 0.0038JN$, $0.057JN$ and $0.10JN$ when $A/\sqrt{2}\sigma = 9\,\mathrm{dB}$, $6\,\mathrm{dB}$ and $5\,\mathrm{dB}$, respectively. Calculating $|x_n|$ then requires $2\bar{N}_c$ real multiplications.

Calculating f_n for $n \in \mathcal{F}$ requires $2N_f$ real multiplications and N_f real divisions. However, N_f is a function of N, which can be see by calculating the mean of N_f as

$$\bar{N}_f = \bar{N}_p \bar{\tau} f_s,$$

where $f_s = \frac{JN}{T}$ is the sampling frequency, and $\bar{N}_p$ is the average number of pulses in an OFDM signal duration, calculated in (4.9). Because $\bar{N}_p$ is the average size of $\mathcal{S}$, we have

$$\bar{N}_f = JNe^{-A^2/2\sigma^2}. \tag{4.53}$$

N_f may change after the first iteration. However, because the OFDM signal after the first iteration is $\hat{x}_n = x_n + c_n$, the N_f for $\hat{x}_n$, denoted as $\hat{N}_f$, is

$$\hat{N}_f = N_f - N_1 + N_2,$$

where N_1 is the number of samples that are higher than A in the first iteration but are lower than A after the first iteration, and N_2 is the number of samples of the newly-generated peaks (i.e., samples that are lower than A in the first iteration but are higher than A after the first iteration).

Because c_n is very small (its average power is usually 10 dB below the OFDM average power, or less), only those peaks of x_n that are slightly smaller or higher than

A will contribute to N_1 or N_2. By using (4.53), N_1 and N_2 can be easily shown to be small numbers, and their difference can be omitted. Thus, N_f is roughly constant in all iterations. We may estimate the execution time of calculating f_n as $2\bar{N}_f$ real multiplications and $\bar{N}_f$ real divisions, or a complexity of $\mathcal{O}(N)$ with a small constant of proportionality. For example, $\bar{N}_f = 3.5 \times 10^{-4}JN$, $0.019JN$ and $0.042JN$ for $A/\sqrt{2}\sigma = 9\,\text{dB}$, $6\,\text{dB}$ and $5\,\text{dB}$, respectively.

In Step 4, finding $\beta^{(\text{opt})}$ requires $4\bar{N}_p$ real multiplications and one real division[5]. From (4.9), $\bar{N}_p$ is a small number proportional to N and is independent of J. Thus, the execution time of this step is determined mainly by the JN-point DFT/IDFT pair and weighting the clipping noise. The latter requires $2JN$ real multiplications for adaptive-scaling and $2N$ real multiplications for constant-scaling.

The fastest way of calculating the DFT/IDFT depends on the number of inputs/outputs of the DFT/IDFT. For example, when A is large and N is small such that the clipping noise usually contains only one or two pulses, direct calculation of DFT/IDFT (having a complexity of $\mathcal{O}(N^2)$ but with small constants of proportionality) may be more preferable. For moderate A and/or large N, FFT algorithms [193–210] must be used. Here, we use the decimation-in-time (DIT) split-radix FFT algorithm [207]. The JN-point DFT in Step 4 has N_f nonzero inputs and N inband outputs (other outputs are not needed). Then its averaged number of real multiplications $\mathcal{M}_{\text{DFT}}$ can be calculated by using (4.79) – (4.81) (see Appendix 4.8) and replacing N_f by $\bar{N}_f$. On the other hand, the JN-point IDFT in Step 4 has N_r nonzero inputs and JN outputs. Its number of real multiplications $\mathcal{M}_{\text{IDFT}}$ can be calculated by using (4.79) – (4.81) and replacing N_f, N and J by N_r, JN and 1.

From the above discussion, the adaptive-scaling algorithm with L iterations requires

$$\mathcal{M}_{\text{AS}} = L(2\bar{N}_f + 2\bar{N}_c + 4\bar{N}_p + \mathcal{M}_{\text{DFT}} + \mathcal{M}_{\text{IDFT}} + 2JN)$$

real multiplications and $L(\bar{N}_f + 1)$ real divisions. Constant-scaling algorithm requires

$$\mathcal{M}_{\text{CS}} = 2\bar{N}_f + 2\bar{N}_c + \mathcal{M}_{\text{DFT}} + \mathcal{M}_{\text{IDFT}} + 2N$$

real multiplications and $\bar{N}_f$ real divisions. The execution time of constant-scaling is smaller than that of adaptive-scaling with one iteration.

[5]Constant-scaling does not need to find a $\beta^{(\text{opt})}$.

The number of iterations L that adaptive-scaling requires to reach a fixed PAR (i.e., independent of N) does not depend on J and N. Based on our analysis, the strength of the clipping pulses and the distances between them are independent of J and N. As well, the adaptive-scaling algorithm proceeds *all* peaks in each iteration. Then, once other parameters (e.g., A, R and the reserved tone set) are fixed, the PAR reduction obtained in each iteration is also fixed no matter what J and N are. Therefore, the complexity of adaptive-scaling for reaching a fixed PAR is $\mathcal{O}(N \log N)$, which is determined by the of complexity of DFT/IDFT. Moreover, because the inputs/outputs of the DFT/IDFT used in our algorithms are sparse, the complexity of DFT/IDFT may be further reduced to $\mathcal{O}(N)$ by using a proper wavelet transform [199, 200].

4.3.3.2 Complexity Analysis of the Active-Set Algorithm

Assume that the active-set algorithm uses the $(2G)$-agon approximation ($G = 4, 8, ...$). Then, the length of the OFDM signal vector and that of the peak-canceling vector proceeded in each iteration is $2GJN$. In each iteration, the execution time of the active-set algorithm is determined by the following steps of this algorithm:

1) Find a suitable optimization direction by solving a set of l linear equations with l variable, where l is the iteration number. We omit the execution time of this step when the total number of iterations L is small.

2) Weight the l peak-canceling kernels associated with the active set. This step requires $lGJN$ real multiplications (because the peak-canceling kernel is symmetric), except in the first iteration.

3) Find the optimal optimization step size μ. If all samples not belonging to the active set are tested, where testing each sample requires one real division, the execution time is approximately $2GJN$ real divisions because the size of the active set is small. The execution time can be reduced by testing only the large peaks of both the OFDM vector and the peak-canceling vector. However, appropriately defining "large peaks" is difficult. Here, we give a lower bound by assuming that the large peaks of the OFDM vector are the peaks higher than A, and that all large peaks of the peak-canceling vector coincide with those of the OFDM vector. Then, the average number of real divisions is $2G\bar{N}_f = 2GJNe^{-A^2/2\sigma^2}$. The execution time

of this step is larger than that of calculating f_n in our algorithms.

4) Weight the peak-canceling signal by μ. This step requires $2GJN$ real multiplications.

With L iterations, the active-set algorithm requires

$$\mathcal{M}_{\text{Act.}} \approx 2LGJN + \sum_{l=2}^{L} lGJN = \frac{1}{2}(L^2 + 5L - 2)GJN$$

real multiplications and $2LGJNe^{-A^2/2\sigma^2}$ real divisions.

The active-set algorithm proceeds one peak (outside the active set) in each iteration. In order to reach a fixed PAR A for different N, all peaks higher than A must be compensated for by the peak-canceling signal. Because the number of peaks higher than A is proportional to N, the required number of iterations is also proportional to N. That is, the complexity of the active-set algorithm for reaching a fixed PAR is[6] $\mathcal{O}(N^3)$.

4.4 Simulation Results

In this section, simulation results are given to verify the estimation of $\bar{\beta}$ used in our constant-scaling algorithm. Then, we compare our proposed algorithms with the active-set algorithm [20] by using simulation. We use $N = 512$, $J = 4$, and 10^6 uniformly distributed 64QAM symbols as the input to the OFDM system. The clipping threshold A is measured in dB with respect to the average signal power before PAR reduction.

4.4.1 Theoretical and Actual Values of β

First, we compare the theoretical value of $\bar{\beta}$ calculated by using (4.34), denoted as $\bar{\beta}_{\text{Theo}}$, and the actual value obtained by using simulation, denoted as $\bar{\beta}_{\text{Simu}}$. Here, $\bar{\beta}_{\text{Simu}}$ is calculated as

$$\bar{\beta}_{\text{Simu}} = E\{\beta_{\text{Simu}}\},$$

where

$$\beta_{\text{Simu}} = \arg\min_{\beta} \sum_{k \in \mathcal{R}} |\hat{F}_k^{(K)} - \beta \hat{F}_k|^2,$$

[6] A similar analysis shows that the complexity of the controlled clipper algorithm [18] is $\mathcal{O}(N^2)$.

and $\hat{F}_k$ and $\hat{F}_k^{(K)}$ are the in-band clipping noise at the first and K-th iterations, respectively.

Fig. 4.5 compares $\bar{\beta}_{\text{Theo}}$ and $\bar{\beta}_{\text{Simu}}$, where the tone-reservation ratio $R = 4.88\%$, 9.96%, 19.92%, respectively; the number of clipping and filtering iterations $K = 20$; and the set of reserved tones $\mathcal{R}$ is randomly selected at the initialization stage. The relative difference d defined as

$$d = \frac{\bar{\beta}_{\text{Theo}} - \bar{\beta}_{\text{Simu}}}{\bar{\beta}_{\text{Simu}}}$$

is shown in Fig. 4.6.

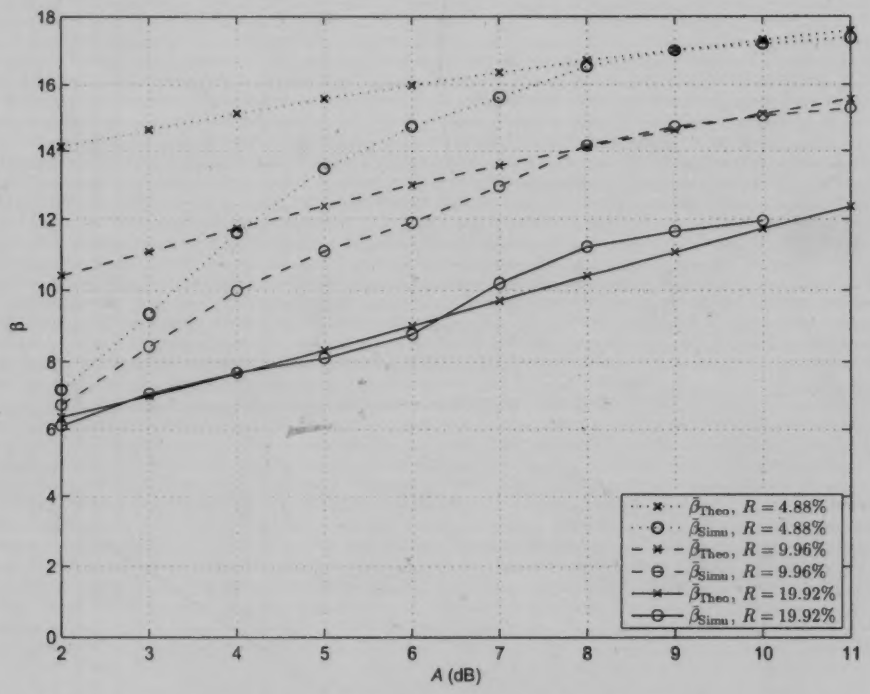


Figure 4.5: Comparison of $\bar{\beta}_{\text{Theo}}$ and $\bar{\beta}_{\text{Simu}}$.

$\bar{\beta}_{\text{Theo}}$ matches $\bar{\beta}_{\text{Simu}}$ when $A \geq 9\,\text{dB}$ because the clipping noise contains only one or two pulses. The difference becomes larger when $A < 9\,\text{dB}$ because of the existence of multiple clipping pulses in the clipping noise. However, the difference may still be acceptable for $A \geq 6\,\text{dB}$ (where $d \leq 10\%$). Therefore, we can use $\bar{\beta}_{\text{Theo}}$ to evaluate the PAR-reduction performance, or use $\bar{\beta}_{\text{Theo}}$ in our constant-scaling algorithm.

The approximation error in $\bar{\beta}_{\text{Theo}}$ depends mainly on the strength of the mainlobe (outside the pulse duration) and the sidelobes of the filtered pulses. This strength is

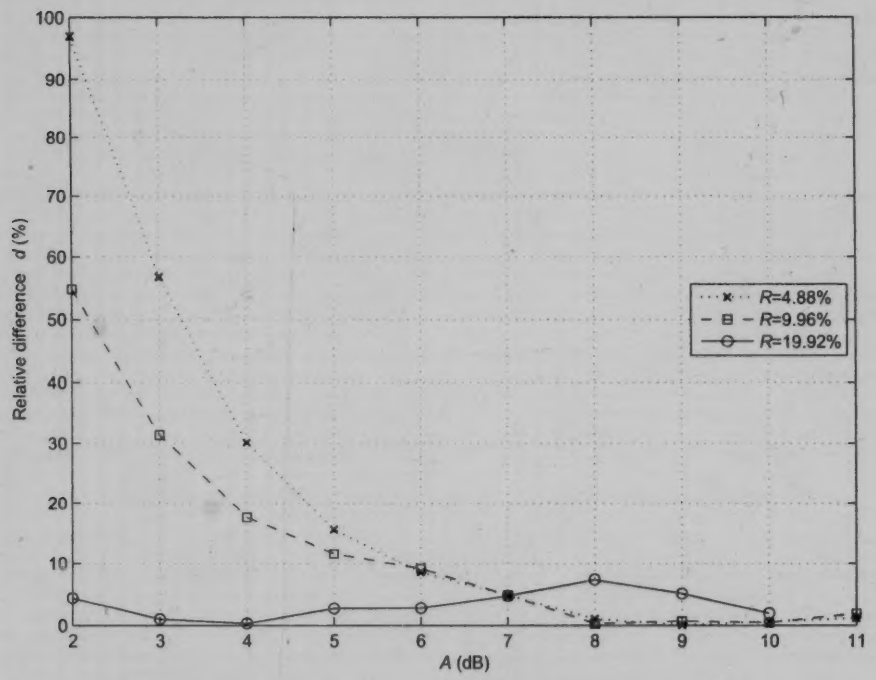


Figure 4.6: Relative difference $\bar{d}$ between $\bar{\beta}_{\text{Theo}}$ and $\bar{\beta}_{\text{Simu}}$.

determined by R and the selection of reserved tones, as well as the distances between the clipping pulses, which is determined by A. When R is large, e.g., $R = 19.92\%$ in Fig. 4.5, the mainlobe outside the pulse duration and the sidelobes of the filtered pulses are small, and the approximation error is small. On the other hand, for small R (e.g., $R = 4.88\%$ or $R = 9.96\%$ in Fig. 4.5), the distances between the clipping pulses are large when A is large (e.g., $A \geq 6\,\text{dB}$ in Fig. 4.5), and the approximation error is small.

Figs. 4.7 and 4.8 illustrate the mean and standard deviation (STD) of the phase and magnitude, respectively, of β_{Simu}, where $N = 512$, $R = 4.88\%$, $A = 6\,\text{dB}$ and $\mathcal{R}$ is randomly selected. Fig. 4.7 shows that both the mean and STD of the phase of β_{Simu} are close to 0. Thus, β_{Simu} can be well approximated as a real value. Fig. 4.8 shows that, for small K, the STD of $|\beta_{\text{Simu}}|$ is small, and $\bar{\beta}_{\text{Theo}}$ closely matches $\bar{\beta}_{\text{Simu}}$. With the increase of K, the STD of $|\beta_{\text{Simu}}|$ and the difference between $\bar{\beta}_{\text{Theo}}$ and $\bar{\beta}_{\text{Simu}}$ become larger.

Fig. 4.9 shows the mean and STD of the normalized approximation error D for

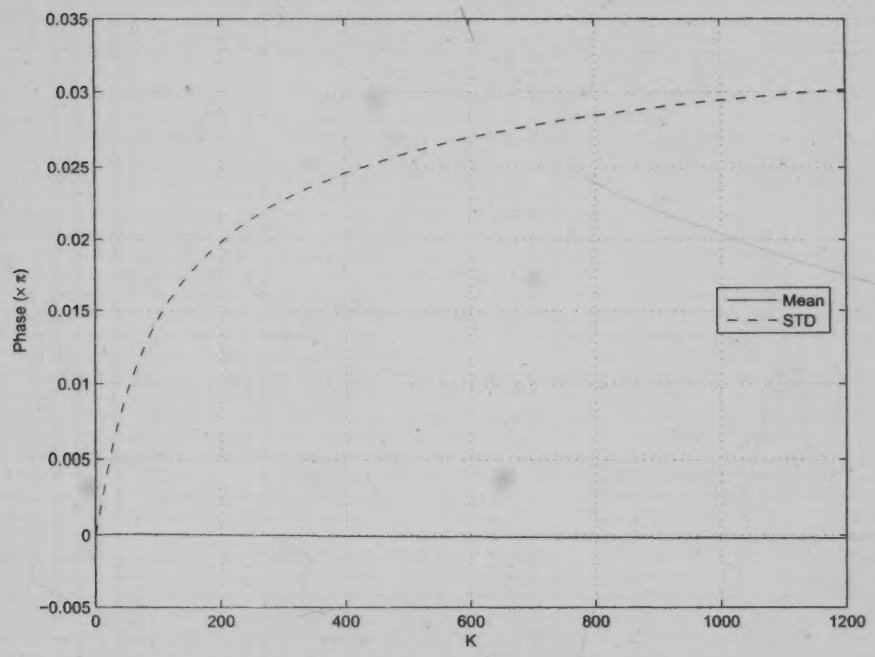


Figure 4.7: Phase of β_{Simu} and its standard deviation.

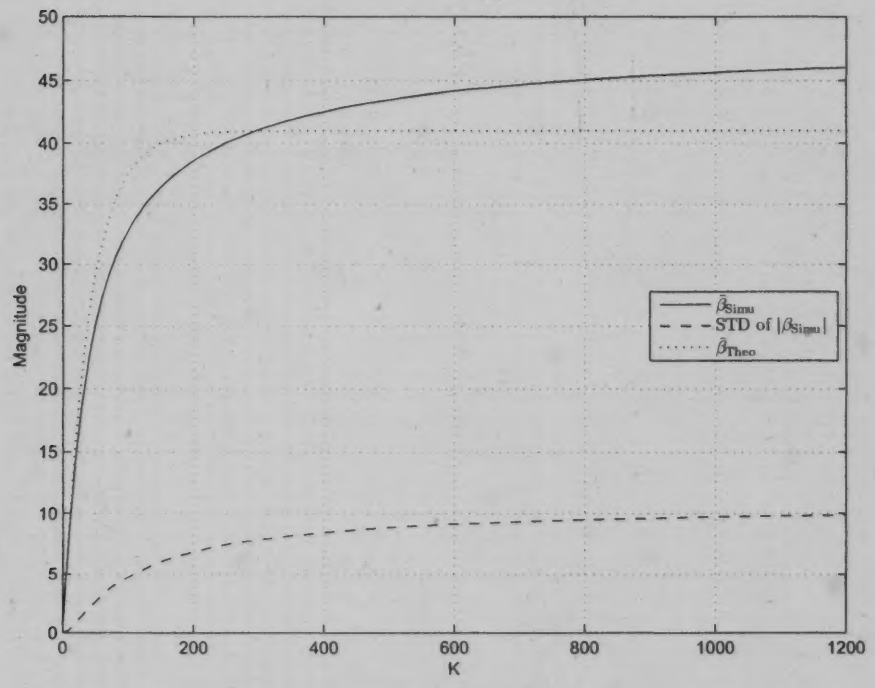


Figure 4.8: Magnitude of β_{Simu} and its standard deviation.

different K when approximating $\hat{F}_k^{(K)}$ by $\beta_{\text{Simu}}\hat{F}_k$. D is defined as

$$D = \frac{\sum_{k\in\mathcal{R}} |\hat{F}_k^{(K)} - \beta_{\text{Simu}}\hat{F}_k|^2}{\sum_{k\in\mathcal{R}} |\hat{F}_k^{(K)}|^2}.$$

The simulation parameters are $N = 512$, $R = 4.88\%$, $A = 6\,\text{dB}$, and $\mathcal{R}$ is randomly selected.

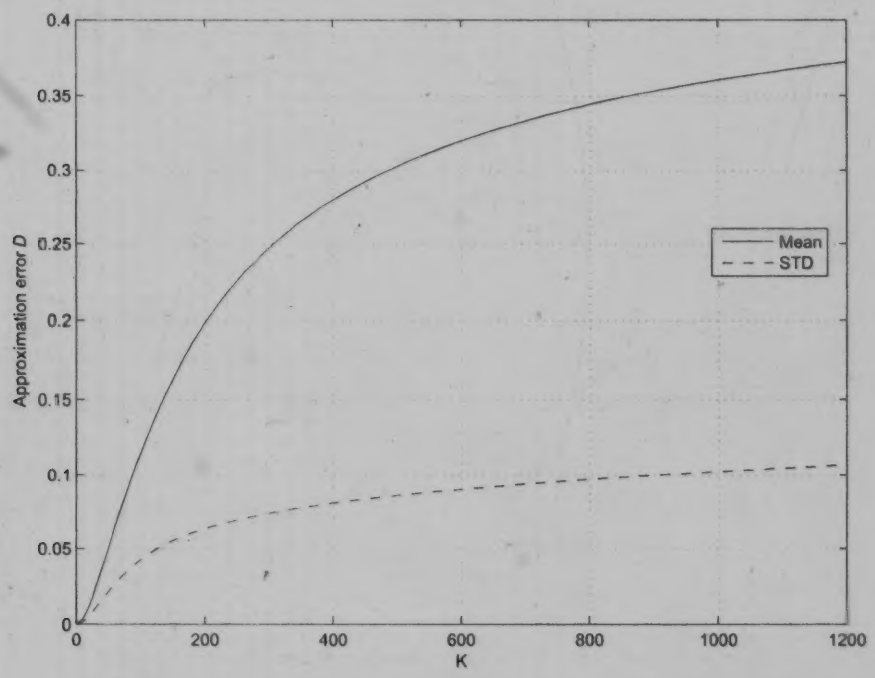


Figure 4.9: Approximation error D and its standard deviation when using β_{Simu}.

We observe that the normalized approximation error D is small when the number of iterations K is small, and D becomes more than 35% when $K \to \infty$.Therefore, K must be carefully selected when using the constant-scaling algorithm because a large K may lead to PAR-reduction performance degradation.

4.4.2 Performance of Proposed Algorithms

We now compare the constant-scaling and adaptive-scaling algorithms with the active-set algorithm under tone-reservation constraints. Here, we use the modified PAR definition (3.18). The PAR CCDF is used to indicate the clip probability. Constant-scaling, where the parameter K is used in constant-scaling to calculate $\bar{\beta}$, has only

one iteration. The execution time of the algorithms we simulated is listed in Table 4.1 in terms of the number of real multiplications/divisions, where L denotes the number of iterations for the adaptive-scaling and the active-set algorithms. As a reference, the execution time of a JN-point FFT is also listed in this table.

Table 4.1: Number of real multiplications/divisions for constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where N=512, $J = 4$.

	$R = 4.88\%$, $A = 6.22\,\text{dB}$	$R = 19.92\%$, $A = 4.96\,\text{dB}$
CS/AS, $L = 1$	22350.68 / 32.16	29273.32 / 89.93
AS, $L = 3$	67052.05 / 96.48	87819.96 / 269.78
AS, $L = 16$	357610.95 / 514.58	468373.12 / 1438.84
Act. Set, $L = 2$	49152.00 / 498.58	49152.00 / 1422.84
Act. Set, $L = 9$	507904.00 / 2243.62	507904.00 / 6402.80
2048-point FFT	16388 / 0	16388 / 0

Fig. 4.10 illustrates the comparison of these algorithms, where the clipping threshold $A = 6.22\,\text{dB}$, the tone-reservation ratio $R = 4.88\%$, and the set of reserved tones $\mathcal{R}$ is randomly selected. The PAR of the original OFDM (no reserved tones) and that of the OFDM with null reserved tones (reserved tones are set to 0) are also plotted in this figure as the dotted and dash-dot curves, respectively.

Setting reserved tones to 0 does not reduce the PAR. For a 10^{-4} clip probability, constant-scaling that approximates 50 clipping and filtering iterations obtains 3.1 dB PAR reduction, which is the same as that of the adaptive-scaling with one iteration, and is about 1.4 dB larger than that of the active-set with two iterations. The adaptive-scaling with three iterations obtains 0.4 dB larger PAR reduction than the active-set with nine iterations for a 10^{-4} clip probability, but with only 13% of the execution time of the latter. With 16 iterations, the adaptive-scaling leads to 5.1 dB PAR reduction (1.3 dB larger than that of the active-set with nine iterations) for a 10^{-4} clip probability. The execution time of this adaptive-scaling is only 70% of that of the active-set with nine iterations.

Fig. 4.11 compares these algorithms for 4.96 dB clipping, 19.92% tone-reservation,

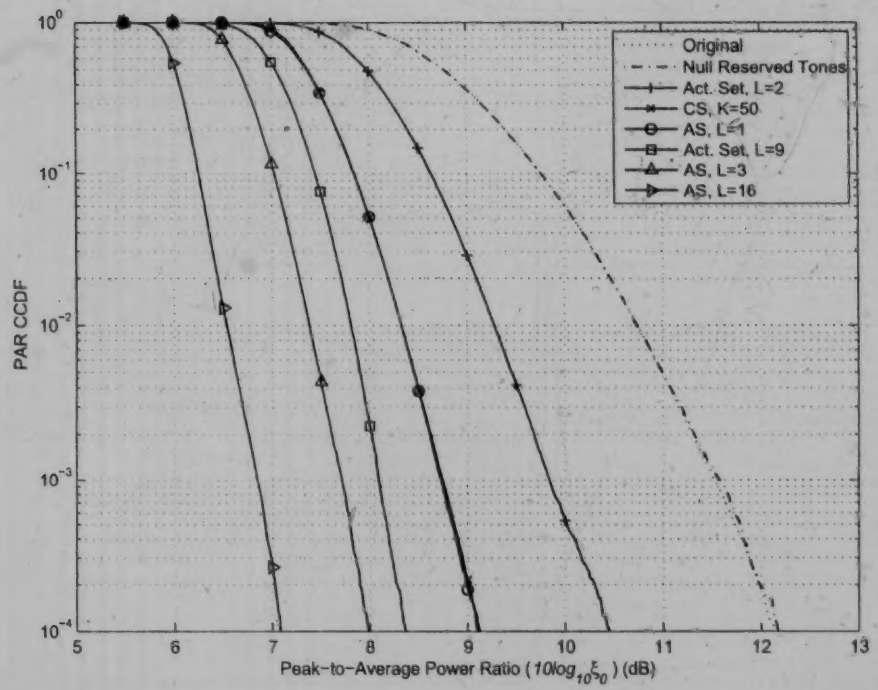


Figure 4.10: PAR reduction of constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where $R = 4.88\%$, $A = 6.22\,\text{dB}$, and $\mathcal{R}$ is randomly selected.

and a randomly selected set of reserved tones. Again, setting the reserved tones to 0 does not reduce the PAR. The adaptive-scaling with one iteration obtains a 4.3 dB PAR reduction, which is about the same as that of the constant-scaling, which approximates 20 clipping and filtering iterations, and is 2.6 dB larger than that of the active-set with two iterations, and 0.7 dB larger than that of the active-set with nine iterations. With three and 16 iterations, adaptive-scaling obtains a 5.7 dB and 6.9 dB reduction (2.1 dB and 3.3 dB larger than that of the active-set with nine iterations) with 17% and 90% of the execution time of the active-set with nine iterations, respectively.

Table 4.2 lists the average power increase in dB. The larger the PAR reduction is, the more the average power increases. However, the largest average power increase is only 0.44 dB. Therefore, the power increase will not significantly increase the BER. Note that the active-set has a negligible average power increase. However, because its PAR reduction is much smaller than that of adaptive-scaling, its BER performance

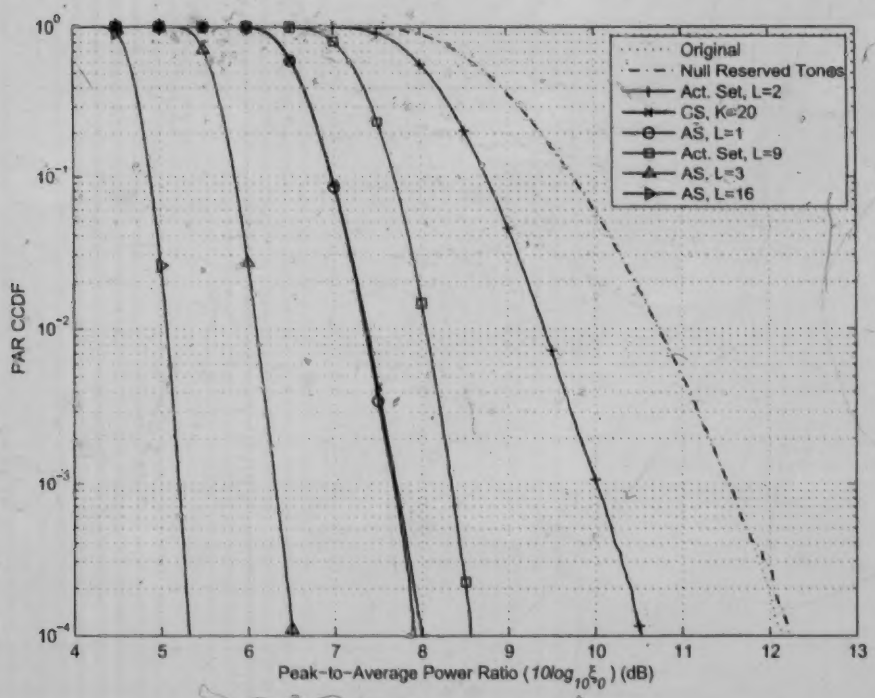


Figure 4.11: PAR reduction of constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where $R = 19.92\%$, $A = 4.96\,\text{dB}$, and $\mathcal{R}$ is randomly selected.

Table 4.2: Average power increase in dB of constant-scaling (CS), adaptive-scaling (AS) and active-set (Act. Set), where the reserved tone set is randomly selected.

	$R = 4.88\%$, $A = 6\,\text{dB}$	$R = 19.92\%$, $A = 4\,\text{dB}$
CS, $K = 50$ for $R = 4.88\%$, $K = 20$ for $R = 19.92\%$	0.14	0.13
AS, $L = 1$	0.11	0.13
AS, $L = 3$	0.23	0.23
AS, $L = 16$	0.44	0.39
Act. Set, $L = 2$	0.02	0.005
Act. Set, $L = 9$	0.11	0.03

is worse than that of the latter.

Given the PAR reduction, execution time and power increase, constant scaling may be a good choice if fast execution is desired. If a large PAR reduction is desired,

the adaptive-scaling with three and 16 iterations, respectively, may be good choices for the 4.88% and 19.92% reserved-tone cases.

Fig. 4.12 compares the BER performance of the adaptive-scaling and the active-set, where 19.92% randomly selected tones are reserved for the PAR reduction, and the clipping threshold is $A = 4.96\,\text{dB}$. The OFDM signal is first processed by using the adaptive-scaling with 16 iterations and the active-set with nine iterations, respectively. The peak reduced signal is passed through a SSPA with a limited linear range. The SSPA output is then transmitted through an AWGN channel. The input/output relationship of SSPA can be written as [1]

$$y(t) = \frac{|x(t)|}{\left(1 + \left(\frac{|x(t)|}{C}\right)^{2p}\right)^{\frac{1}{2p}}} e^{j\phi(t)},$$

where $x(t) = |x(t)|e^{j\phi(t)}$ is the input, and $y(t)$ is the output of SSPA. Usually, $p = 3$ for practical SSPA. In our simulation, the saturation point $C = 5.46\,\text{dB}$. In Fig. 4.12, the BER of the original OFDM signal (without PAR reduction) passing through the SSPA with $C = 5.46\,\text{dB}$ and passing through an ideal power amplifier with infinite linear range, respectively, are also included for reference.

If an ideal power amplifier is used, the OFDM system has a BER of 10^{-6} when $E_b/N_0 = 18.6\,\text{dB}$. However, if the SSPA with $C = 5.46\,\text{dB}$ is used, the OFDM system has a BER floor of 6×10^{-3}. By using the adaptive-scaling with 16 iterations, the OFDM system obtains a BER of 10^{-6} with $E_b/N_0 = 22.0\,\text{dB}$. On the other hand, the active-set with nine iterations has a BER floor of 2×10^{-5}.

Fig. 4.13 compares the radiation out of the OFDM frequency band $\left[-\frac{N}{2T}, \frac{N}{2T}\right]$ for the adaptive-scaling and active-set. The simulation parameters are the same as above. If no PAR reduction is used, the out-of-band radiation is $-24.5\,\text{dB}$. By using the active-set with nine iterations, the out-of-band radiation is reduced to $-29.5\,\text{dB}$. However, by using the adaptive-scaling with 16 iterations, the out-of-band radiation is only $-34.5\,\text{dB}$.

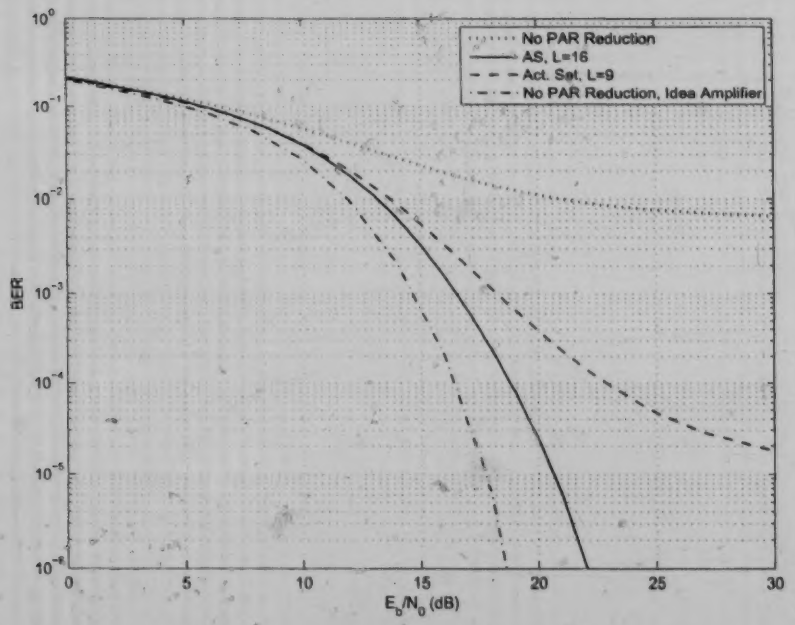


Figure 4.12: BER comparison of adaptive-scaling (AS) and active-set (Act. Set), where $R = 19.92\%$, $A = 4.96\,\text{dB}$, $\mathcal{R}$ is randomly selected, and $C = 5.46\,\text{dB}$.

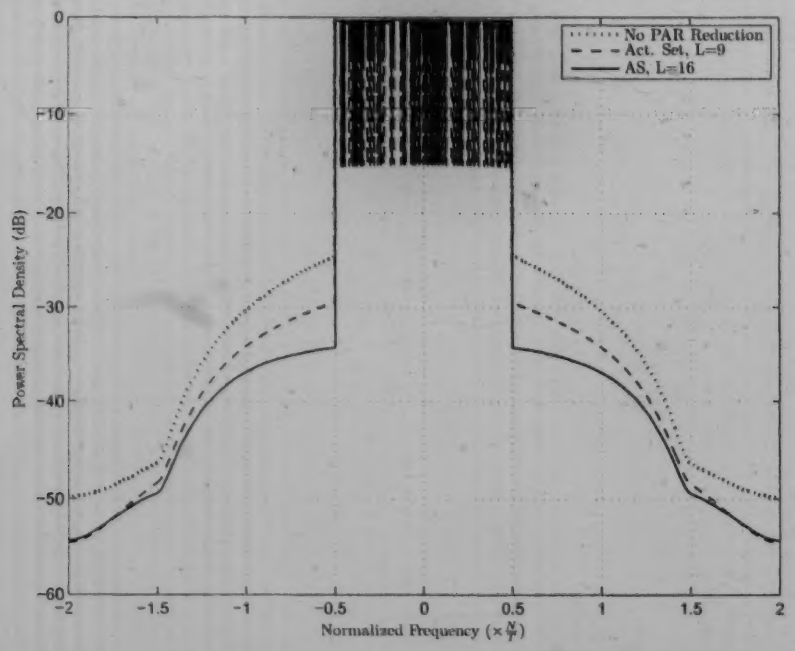


Figure 4.13: Out-of-band radiation of adaptive-scaling (AS) and active-set (Act. Set), where $R = 19.92\%$, $A = 4.96\,\text{dB}$, $\mathcal{R}$ is randomly selected, and $C = 5.46\,\text{dB}$.

4.4.3 Tradeoff of PAR Reduction and Execution Time for the Adaptive-Scaling Algorithm

Tradeoffs for the adaptive-scaling algorithm can be made between the PAR reduction and execution time. In this subsection, we find the optimal number of iterations by using simulation.

Fig. 4.14 illustrates the PAR-reduction performance of the adaptive-scaling algorithm at the first 16 iterations, where the clipping threshold is 4.96 dB, the tone-reservation ratio is 19.92%, and the reserved tones are randomly selected. In this figure, the curves represent the PAR distribution for the adaptive-scaling with, from right to left, one to 16 iterations.

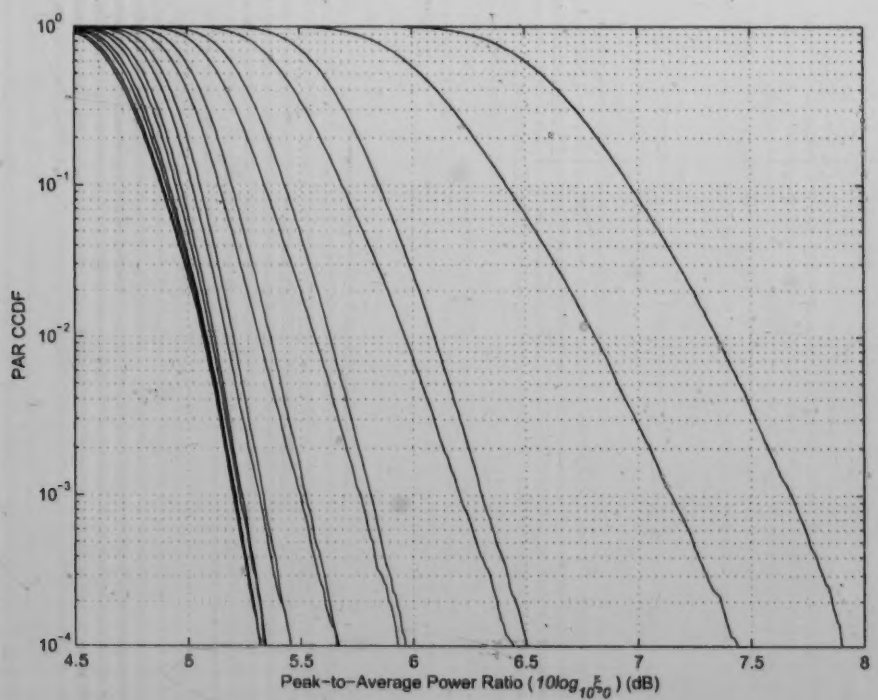


Figure 4.14: PAR reduction of adaptive-scaling at the first 16 iterations, where $R = 19.92\%$, $A = 4.96\,\text{dB}$, and $\mathcal{R}$ is randomly selected.

The $(2l+1)$-th iteration of the adaptive-scaling exhibits approximately the same PAR-reduction performance as the $2l$-th iteration for $l \geq 1$. The adaptive-scaling with three iterations may be a good choice to balance PAR reduction and execution time. On the other hand, the adaptive-scaling with five iterations reduces the PAR

to about 6 dB at a 10^{-4} clip probability. The next 11 iterations further obtain only a 0.7 dB PAR reduction at a 10^{-4} clip probability.

Similarly, Fig. 4.15 shows that, when a 6.22 dB clipping threshold and a 4.88% randomly selected reserved tone set are used, three iterations may be a good tradeoff in order for adaptive-scaling to balance PAR reduction and execution time. The next 13 iterations further obtain only a 0.9 dB PAR reduction at a 10^{-4} clip probability.

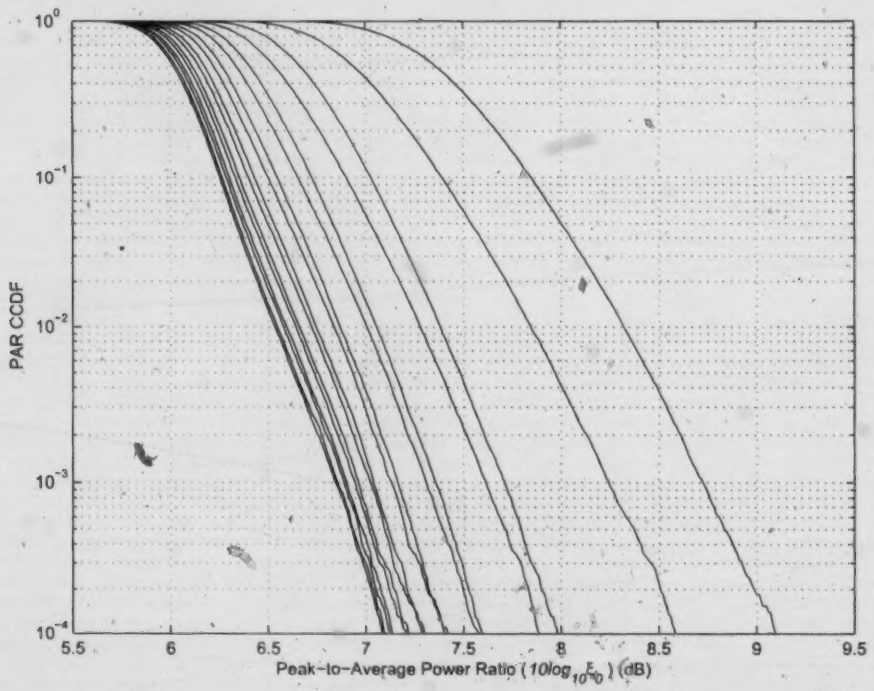


Figure 4.15: PAR reduction of adaptive-scaling at the first 16 iterations, where $R = 4.88\%$, $A = 6.22\,\text{dB}$, and $\mathcal{R}$ is randomly selected.

4.5 The Constant-Scaling Algorithm: when Inband Distortion Is Allowed

The nonlinearity of the HPA introduces both inband distortion, which increases the BER, and out-of-band radiation, which interferes with neighboring communications. While the latter is unacceptable, the former may be tolerable in low bit rate communications. Thus, we may apply the constant-scaling algorithm, as a substitution

for the iterative clipping and filtering algorithm [90], to suppress the out-of-band radiation [19].

In distortion-tolerable cases, the clipping noise is distributed on all the subcarriers and mixed with the data symbols. In terms of PAR reduction, this situation can be considered as a special case of tone-reservation where $R = 1$, and the constraint of (3.17) is discarded. The above discussion leads to the following conclusions:

1. The constraint of (3.17) affects the BER performance and the average power. However, this constraint does not affect the performance of the peak reduction.

2. Applying (3.17) in PAR reduction ensures the orthogonality of the data symbols and the clipping noise, but slightly increases the average power. The BER is slightly increased only due to the increased average power.

3. Discarding (3.17) in PAR reduction decreases the average power. However, the BER is increased due to the clipping noise. However, the clipping noise can be partly canceled at the receiver by using clipping noise cancelation algorithms [86, 88, 89, 91, 95].

Letting $R = 1$ and discarding (3.17), we apply the constant-scaling algorithm that approximates three clipping and filtering iterations to an OFDM system with 256 subcarriers and a unitary QPSK symbol input. We compare the constant-scaling algorithm to the iterative clipping and filtering algorithm with three iterations in terms of PAR reduction, BER and out-of-band radiation. In our simulations, we consider the 6 dB and 3 dB clipping cases. The clipping threshold A is small in the 3 dB clipping case. However, our simulation results show that the constant-scaling algorithm can also be used together with clipping noise cancelation techniques [86,88].

Fig. 4.16 compares the PAR reduction of the constant-scaling algorithm and that of the iterative clipping and filtering (ICF) algorithm. Their PAR reduction is close. For the 6dB clipping case, our proposed technique is about 0.3dB better than the iterative clipping and filtering at a 10^{-4} clip probability.

To investigate BER performance degradation and out-of-band distortion, we consider passing the PAR-reduced signal through an SSPA with limited linear range. The input/output relationship of SSPA is given in (3.3).

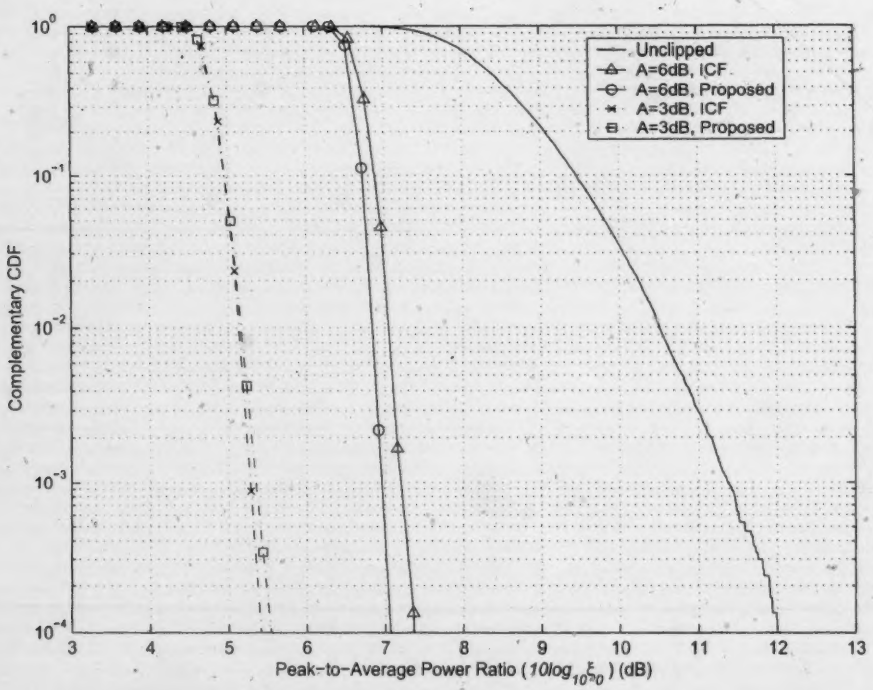


Figure 4.16: PAR CCDF comparison of the constant-scaling algorithm and the iterative clipping and filtering algorithm.

For the 6 dB clipping case, the clipping threshold of both algorithms is $A = 6$ dB. After peak reduction, the PAR-reduced signal is passed through the SSPA with $C = 6$ dB. For the 3 dB clipping, both A and C are set to 3 dB. Figures 4.17 compares the BER performance of both algorithms. As a reference, the dotted curve shows the BER of the ideal case when the HPA has enough linear range and no PAR-reduction technique is needed. The solid curves represent the BER when using the clipping noise cancelation technique [86] while the dashed curves represent the BER without clipping noise cancelation. The "$\circ$" is for the constant-scaling algorithm, and the "$\times$" is for the iterative clipping and filtering algorithm.

For the 6 dB clipping, the BER performances of these two algorithms are very close, and there is little need to use clipping noise cancelation. The BER performance in this case is only about 1 dB worse than that of the ideal case. On the other hand, in the 3 dB clipping both techniques are quite worse than that of the ideal case (6.5 – 8.5 dB at the BER of 10^{-5}) when no clipping noise cancelation is used, while the

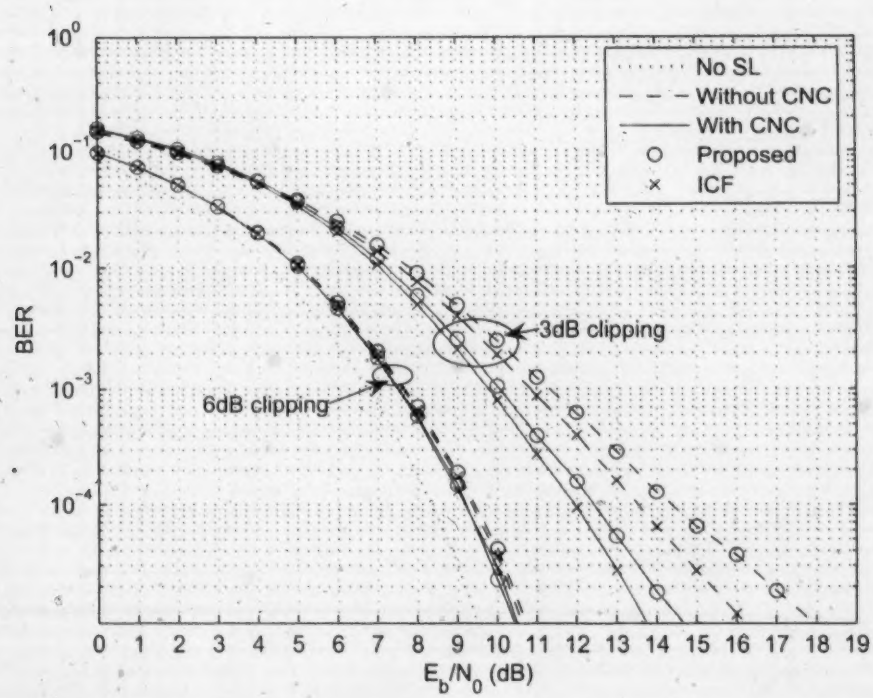


Figure 4.17: BER performance comparison of the constant-scaling algorithm and the iterative clipping and filtering algorithm.

gap between these two techniques at the BER of 10^{-5} is about 2 dB. However, when clipping noise cancelation is used, the BER performance of these two techniques is improved by 2.5 – 3.5 dB, and the constant-scaling algorithm is only about 1 dB worse than the iterative clipping and filtering algorithm. Therefore, the performance of the constant-scaling algorithm is comparable to that of the iterative clipping and filtering algorithm. However, because the constant-scaling algorithm performs only one iteration, its execution time is significantly less than that of iterative clipping and filtering at both the transmitter and the receiver.[7]

The out-of-band radiation comparison for the 3 dB clipping case is shown in Fig. 4.18. Here, pulse shaping is not considered for simplicity. Fig. 4.18 shows that iterative clipping and filtering leads to about 6 dB lower out-of-band radiation than that obtained without using any PAR-reduction technique, and our proposed technique leads to about 2.5 dB lower out-of-band radiation than that obtained by

[7]As in [86], the same number of iterations has to be performed at the receiver to cancel the clipping noise.

iterative clipping and filtering.

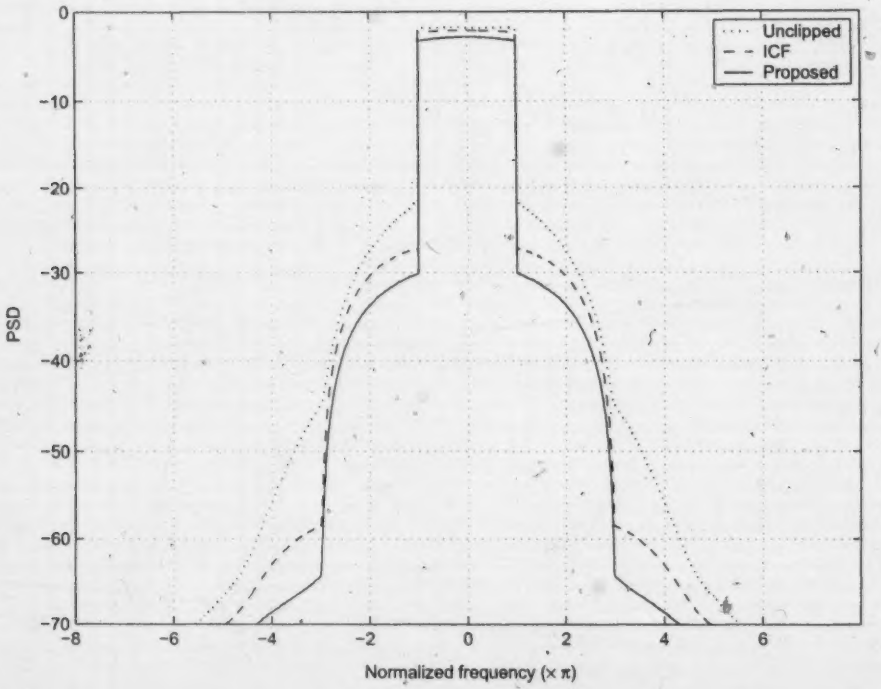


Figure 4.18: Out-of-band radiation comparison of the constant-scaling algorithm and the iterative clipping and filtering algorithm for the 3 dB clipping case.

4.6 Conclusions

In this chapter, using a parabolic approximation of clipping pulses, we analyzed the peak regrowth and the flat power spectrum of the in-band clipping noise of tone-reservation by using iterative clipping and filtering. We showed that the clipping noise obtained after several clipping and filtering iterations is approximately proportional to that generated in the first iteration, and derived the constant of proportionality by using the level-crossing theory. We have also proposed a constant scaling algorithm and an adaptive scaling algorithm for tone-reservation. These algorithms scale the filtered clipping noise by a constant or an adaptively calculated factor to generate a peak-canceling signal. Our simulation showed that the PAR and the execution time of the proposed algorithms are lower than those of the active-set algorithm.

4.7 Appendices

4.7.A Time Independence of Different Clipping Pulses

$x(t)$ can be shown to be a cyclostationary process. Then, the correlation of $x(t)$ of time t_i and t_k depends only on the time difference $\Delta t = t_i - t_k$. The correlation of $x(t)$ and $x(t+\Delta t)$ is

$$\begin{aligned} R_x(\Delta t) &= \frac{1}{2}E\{x(t)x^*(t+\Delta t)\} \\ &= \frac{1}{2N}\sum_{m=-N/2}^{N/2-1}\sum_{l=-N/2}^{N/2-1} E\{X_m X_l^*\}e^{j2\pi(mt_i - l(t+\Delta t))/T} \\ &= \frac{\sigma^2}{N}\sum_{m=-N/2}^{N/2-1} e^{j2\pi m\Delta t/T} \\ &= \sigma^2\frac{\sin(\pi N\Delta t/T)}{N\sin(\pi\Delta t/T)}e^{-j\pi\Delta t/T}. \end{aligned} \tag{4.54}$$

Because $\frac{1}{2}E\{x(t)x(t)^*\} = \sigma^2$, the correlation coefficient of $x(t)$ and $x(t+\Delta t)$ is

$$\rho_x(\Delta t) = \frac{\sin(\pi N\Delta t/T)}{N\sin(\pi\Delta t/T)}e^{-j\pi\Delta t/T}. \tag{4.55}$$

Strictly speaking, $\rho_x(n) = \delta(n)$, and the samples of $x(n)$ are independent only when the Nyquist sampling rate T/N is used. However, we will show that, in the continuous-time domain, the possibility that two or more clipping pulses fall within a small time interval and have a large correlation is small and can be omitted. Then, the clipping pulses occurring at different time instances can be effectively treated as independent.

Without loss of generality, let $\Delta t = nT/N$, where n is a real number, and $0 \le n < N$. When n is small[8] compared to N, then,

$$|\rho_x(nT/N)| = \left|\frac{\sin(\pi n)}{N\sin(\pi n/N)}\right| = \left|\frac{\sin(\pi n)}{\pi n}\right|, \quad (\text{when } N\to\infty). \tag{4.56}$$

For example, $|\rho_x(nT/N)| \approx 0.071$ when $n = 4.5$.

The probability that two or more clipping pulses occur within a time interval of $4.5T/N$ is small. [190] shows that the up-crossing time of $x(t)$ is Poisson distributed when $N \to \infty$:

$$\lim_{A\to\infty}\Pr[U_{n,A}(0,\Delta t) = k] = \frac{(\Delta t\lambda_A)^k e^{-\Delta t\lambda_A}}{k!}, \tag{4.57}$$

[8]We do not need to consider a large n because, at a large n, $\rho_x(n)$ is virtually 0.

where $U_{n,A}(0, \Delta t)$ is the number of the up-crossing that a χ^2 random process

$$Y(t) = X_1^2(t) + X_2^2(t) + \cdots + X_n^2(t) \tag{4.58}$$

up-crosses a fixed level $A/\sigma > 0$, with $\sigma = E\{X_k^2(t)\}$ for all k, during the time interval $(0, \Delta t)$, and λ_A is the up-crossing rate. Based on our assumption of a large A, each up-crossing corresponds to a parabolic clipping pulse. Then, given the condition that a clipping pulse occurs in $(0, \Delta t)$, the probability that more than m ($m \geq 1$) clipping pulses occur in the same time interval $(0, \Delta t)$, denoted as $\Pr(m)$, is

$$\begin{aligned} \Pr(m) &= \Pr[U_{n,A}(0, \Delta t) > m | \text{A clipping pulse occurs}] \\ &= \Pr[U_{n,A}(0, \Delta t) > m - 1] \\ &= 1 - \sum_{l=0}^{m-1} \frac{(\Delta t \lambda_A)^l e^{-\Delta t \lambda_A}}{l!}. \end{aligned} \tag{4.59}$$

By substituting $\Delta t = nT/N$ into (4.59), $\Pr(m)$ is independent of N. Table 4.3 lists $\Pr(m)$ for different clipping thresholds, where $n = 4.5T/N$. This table shows that, when A is large, the chance that two or more clipping pulses will fall within the same time interval $(0, 4.5T/N)$ is small. In other words, most clipping pulses are a large "distance" apart from each other, so that the $|\rho_x(\Delta t)|$'s between these pulses are small and can be approximated as 0. Because $x(t)$ is Gaussian, $x(t)$ and $x(t + \Delta t)$ are also independent with respect to Δt.

Table 4.3: Given that a clipping pulse occurs in $(0, 4.5T/N)$, the probability that more than k ($k \geq 1$) clipping pulses occur in the same time interval $(0, 4.5T/N)$.

A(dB)	3	6	9
Pr(1)	0.59	0.16	4.6×10^{-3}
Pr(2)	0.22	1.3×10^{-2}	1.1×10^{-5}
Pr(3)	6.0×10^{-2}	7.4×10^{-4}	1.6×10^{-8}
Pr(4)	1.3×10^{-2}	3.1×10^{-5}	1.9×10^{-11}
Pr(5)	2.2×10^{-3}	1.1×10^{-6}	1.7×10^{-14}

Table 4.3 also includes a low clipping threshold case, where $A = 3\,\text{dB}$. In this case, Pr(1) and Pr(2) are relatively large at $n = 4.5$. However, taking into account that the $|\rho_x(\Delta t)|$ is only 0.071 at $n = 4.5$, we may still treat the clipping pulses in this case as uncorrelated to simplify the power spectrum estimation.

4.7.B The Conditional pdf and moments of γ_i

In this section, we find the conditional pdf and moments of $\gamma_i = \dfrac{\dot{x}_I(t_i)}{|x_R(t_i)|}$, given $\dot{r}(t_i) = 0$ and $r(t_i) \geq A$. For the ease of notation, we drop off the subscript i and the time index t_i in the following analysis. Strictly speaking, we also need a condition of $\ddot{r} \leq 0$. However, Appendix 4.7.C shows that $\Pr[\ddot{r}|\dot{r} = 0, r \leq A] \approx 0$ unless A is small.

First, we show that x_R, x_I, $\dot{x}_R$ and $\dot{x}_I$ are independent when N is large. $\dot{x}_R = \dfrac{1}{2}(\dot{x} + \dot{x}^*)$ and $\dot{x}_I = \dfrac{1}{2j}(\dot{x} - \dot{x}^*)$, where

$$\dot{x}(t) = \frac{dx(t)}{dt} = \frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} \frac{j2\pi k}{T} X_k e^{j2\pi kt/T}. \tag{4.60}$$

As well, $E\{X_k X_l\} = 0$ for any k and l. Then, $\dot{x}_R$ and $\dot{x}_I$ are i.i.d. Gaussian processes with zero mean and variance

$$\dot{\sigma}^2 = \frac{4\pi^2\sigma^2}{NT^2} \sum_{k=-N/2}^{n/2-1} k^2 = \frac{(N^2+2)\pi^2\sigma^2}{3T^2}.$$

When N is large, $\dot{\sigma}^2 \approx \dfrac{\pi^2 N^2 \sigma^2}{3T^2} = \dfrac{\pi^2}{3} W^2 \sigma^2$, which agrees with (4.7).

Using (4.1) and (4.60), we have

$$E\{x_R \dot{x}_R\} = E\{x_I \dot{x}_I\} = 0.$$

On the other hand,

$$E\{x_R \dot{x}_I\} = -E\{x_I \dot{x}_R\} = \frac{2\pi\sigma^2}{NT} \sum_{k=-N/2}^{N/2-1} k = -\frac{\pi\sigma^2}{T},$$

and their correlation coefficients are

$$\rho_{x_R \dot{x}_I} = -\rho_{x_I \dot{x}_R} = -\frac{3}{\sqrt{N^2+2}},$$

which is zero when $N \to \infty$ (less than 0.014 when $N \geq 128$). Therefore, x_R, x_I, $\dot{x}_R$ and $\dot{x}_I$ are independent when N is large.

We now find the joint pdf $p(\dot{x}_I, x_R, x_I|\dot{r} = 0)$. Because x_R, x_I, $\dot{x}_R$ and $\dot{x}_I$ are independent, fixing x_R and x_I does not change the distribution of $\dot{x}_R$ and $\dot{x}_I$. Note that

$$\dot{r} = \frac{1}{r}(x_R \dot{x}_R + x_I \dot{x}_I).$$

Then, given x_R and x_I, $\dot{r}$ is also a Gaussian process with zero mean and variance

$$\dot{\sigma}_r^2 = E\{\dot{r}^2\} = \dot{\sigma}^2.$$

Because $\dot{\sigma}_r^2$ is independent of x_R and x_I, $\dot{r}$ is independent of x_R, x_I and r, and

$$p(\dot{r}) = \frac{1}{\sqrt{2\pi}\dot{\sigma}} e^{-\dot{r}/2\dot{\sigma}^2}.$$

Given x_R and x_I, the correlation coefficient between $\dot{x}_I$ and $\dot{r}$ is

$$\rho_{\dot{x}_I,\dot{r}} = \frac{x_I}{r}.$$

Then,

$$p(\dot{x}_I, \dot{r}|x_R, x_I) = \frac{\sqrt{x_R^2 + x_I^2}}{2\pi\dot{\sigma}^2|x_R|} \exp\left(-\frac{x_R^2 + x_I^2}{2\dot{\sigma}^2 x_R^2}(\dot{x}_I^2 - 2\frac{x_I\dot{x}_I\dot{r}}{\sqrt{x_R^2 + x_I^2}} + \dot{r}^2)\right), \tag{4.61}$$

and

$$\begin{aligned} p(\dot{x}_I, x_R, x_I|\dot{r} = 0) &= \left.\frac{p(\dot{x}_I, \dot{r}|x_R, x_I)p(x_R)p(x_I)}{p(\dot{r})}\right|_{\dot{r}=0} \\ &= \frac{\sqrt{x_R^2 + x_I^2}}{(2\pi)^{3/2}\dot{\sigma}\sigma^2|x_R|} \exp\left(-\frac{(x_R^2 + x_I^2)\dot{x}_I^2}{2\dot{\sigma}^2 x_R^2} - \frac{x_R^2 + x_I^2}{2\sigma^2}\right). \end{aligned} \tag{4.62}$$

We use the following transforms to obtain $p(\gamma, r, \theta|\dot{r} = 0)$:

$$\dot{x}_I = r\gamma|\cos\theta| + \dot{r}\sin\theta, \tag{4.63}$$

$$x_R = r\cos\theta, \tag{4.64}$$

$$x_I = r\sin\theta. \tag{4.65}$$

The Jacobian of the transformation from $\dot{x}_I$, x_R, x_I to γ, r and θ is

$$\mathfrak{J} = \left.\begin{vmatrix} r\cos\theta & \gamma\cos\theta & -r\gamma\sin\theta + \dot{r}\cos\theta \\ 0 & \cos\theta & -r\sin\theta \\ 0 & \sin\theta & r\cos\theta \end{vmatrix}\right|_{\dot{r}=0} = r^2\cos\theta,$$

for $-\pi/2 \le \theta \le \pi/2$, and

$$\mathfrak{J} = \left.\begin{vmatrix} -r\cos\theta & -\gamma\cos\theta & r\gamma\sin\theta + \dot{r}\cos\theta \\ 0 & \cos\theta & -r\sin\theta \\ 0 & \sin\theta & r\cos\theta \end{vmatrix}\right|_{\dot{r}=0} = -r^2\cos\theta,$$

for $-\pi/2 \leq \theta \leq \pi/2$. Thus,

$$\mathfrak{J} = r^2|\cos\theta| \tag{4.66}$$

for $0 \leq \theta \leq 2\pi$. Then, we have

$$\begin{aligned} p(\gamma, r, \theta|\dot{r} = 0) &= p(\dot{x}_I, x_R, x_I|\dot{r} = 0)|\mathfrak{J}| \\ &= \frac{r^2}{(2\pi)^{3/2}\dot{\sigma}\sigma^2} \exp\left(-\frac{r^2\gamma^2}{2\dot{\sigma}^2} - \frac{r^2}{2\sigma^2}\right). \end{aligned} \tag{4.67}$$

The pdf of γ conditioned on $\dot{r} = 0$ and $r \geq A$ is

$$\begin{aligned} p(\gamma|\dot{r} = 0, r \geq A) &= \frac{\int_A^\infty \int_0^{2\pi} p(\gamma, r, \theta|\dot{r} = 0) d\theta dr}{\int_A^\infty p(r) dr} \\ &= \frac{A^3}{4\sqrt{2\pi}\dot{\sigma}\sigma^2\psi^3} \left(2\psi e^{-\psi^2} + \sqrt{\pi}\mathrm{erfc}\,(\psi)\right) e^{A^2/2\sigma^2}, \end{aligned} \tag{4.68}$$

where

$$\psi = \frac{A\sqrt{\gamma^2\sigma^2 + \dot{\sigma}^2}}{\sqrt{2}\dot{\sigma}\sigma}.$$

The conditional cumulative distribution function (cdf) of γ cannot be written in a closed form. However, because $p(\gamma|\dot{r} = 0, r \geq A) = p(-\gamma|\dot{r} = 0, r \geq A)$, the conditional mean $m_\gamma = E\{\gamma|\dot{r} = 0, r \geq A\} = 0$. The conditional variance $\sigma_\gamma^2 = E\{\gamma^2|\dot{r} = 0, r \geq A\}$ can be found by using (4.62).

We first transform $p(\dot{x}_I, x_R, x_I|\dot{r} = 0)$ to $p(\dot{x}_I, r, \theta|\dot{r} = 0)$. The Jacobian of the transformation is $\mathfrak{J} = r$. Then,

$$\begin{aligned} \sigma_\gamma^2 &= \frac{\int_{-\infty}^\infty \int_A^\infty \int_0^{2\pi} \frac{\dot{x}_I^2}{r^2\cos^2\theta} p(\dot{x}_I, r, \theta|\dot{r} = 0) d\theta dr d\dot{x}_I}{\int_{-\infty}^\infty \int_A^\infty \int_0^{2\pi} p(\dot{x}_I, r, \theta|\dot{r} = 0) d\theta dr d\dot{x}_I} \\ &= \frac{\pi^2 W^2}{6} \mathrm{E}_1\left(\frac{A^2}{2\sigma^2}\right) e^{A^2/2\sigma^2}, \end{aligned} \tag{4.69}$$

and

$$\sigma_\gamma^4 = \frac{\pi^4 W^4}{12} \left(\mathrm{E}_1\left(\frac{A^2}{2\sigma^2}\right) e^{A^2/2\sigma^2} - \frac{2\sigma^2}{A^2}\right). \tag{4.70}$$

We may also find the central moments of $|\gamma|$ by using (4.62). The conditional mean of $|\gamma|$ is

$$\begin{aligned} m_{|\gamma|} &= E\left\{|\gamma||\dot{r} = 0, r \geq A\right\} \\ &= \frac{\int_{-\infty}^\infty \int_A^\infty \int_0^{2\pi} \frac{|\dot{x}_I|}{r|\cos\theta|} p(\dot{x}_I, r, \theta|\dot{r} = 0) d\theta dr d\dot{x}_I}{\int_{-\infty}^\infty \int_A^\infty \int_0^{2\pi} p(\dot{x}_I, r, \theta|\dot{r} = 0) d\theta dr d\dot{x}_I} \\ &= \frac{\pi W}{\sqrt{3}} \mathrm{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right) e^{A^2/2\sigma^2}. \end{aligned} \tag{4.71}$$

The conditional variance of $|\gamma|$ is $\sigma^2_{|\gamma|} = \sigma^2_\gamma - m^2_{|\gamma|}$.

When $A \to \infty$, $m_{|\gamma|}$ and $\sigma_{|\gamma|}$ are virtually zero. For practical A's, $m_{|\gamma|}$ and $\sigma_{|\gamma|}$ are also small compared to the OFDM half bandwidth (in rad/s) πW. Table 4.4 lists the theoretical and simulated values of $m_{|\gamma|}$ and $\sigma^2_{|\gamma|}$ for different A. γ_i introduces a frequency shift to $F_i(\omega)$, the frequency spectrum of the clipping pulse $f_i(t)$. Moreover, $F_i(\omega)$ reaches its maximum magnitude at $\omega = 0$ if $\gamma_i = 0$. Then, we can measure the frequency at which $F_i(\omega)$ reaches its maximum magnitude to obtain the simulated statistics of γ_i. In our simulation, we use $N = 512$, $J = 128$ and QPSK input symbols[9].

Table 4.4: $m_{|\gamma|}$ and $\sigma_{|\gamma|}$ for different A. Simulated results are obtained with $N = 512$, $J = 128$ and QPSK input symbols.

$A/\sqrt{2}\sigma$(dB)	3	6	9
$m_{|\gamma|}$, Theoretical ($\times\pi W$)	0.19	0.15	0.11
$m_{|\gamma|}$, Simulated ($\times\pi W$)	0.16	0.13	0.11
$\sigma_{|\gamma|}$, Theoretical ($\times\pi W$)	0.15	0.11	0.083
$\sigma_{|\gamma|}$, Simulated ($\times\pi W$)	0.12	0.10	0.078

4.7.C $\Pr[\ddot{r}(t_i) > 0 | \dot{r}(t_i) = 0, r(t_i) \geq A] \to 0$ when $A \to \infty$

In this section, we prove that $\Pr[\ddot{r}(t_i) > 0 | \dot{r}(t_i) = 0, r(t_i) \geq A] \to 0$ when $A \to \infty$. In the following, we drop off the time index t_i for the ease of notation.

Because $\ddot{x}_R = \frac{1}{2}(\ddot{x} + \ddot{x}^*)$ and $\ddot{x}_I = \frac{1}{2j}(\ddot{x} - \ddot{x}^*)$, where

$$\ddot{x}(t) = \frac{d\dot{x}(t)}{dt} = -\frac{1}{\sqrt{N}} \sum_{k=-N/2}^{N/2-1} \frac{4\pi^2 k^2}{T^2} X_k e^{j2\pi kt/T}, \tag{4.72}$$

$\ddot{x}_R$ and $\ddot{x}_I$ are i.i.d. Gaussian processes with zero mean and variance[10]

$$\ddot{\sigma}^2 = \frac{16\pi^4\sigma^2}{NT^4} \sum_{k=-N/2}^{N/2-1} k^4 = \frac{\pi^4\sigma^2}{15T^4}(3N^4 + 20N^2 - 8)$$
$$\approx \frac{N^4\pi^4\sigma^2}{5T^4} = \frac{\pi^4}{5} W^4\sigma^2.$$

[9]We choose a large oversampling factor J to avoid the case in which the clipping pulse has only one nonzero sample.

[10]We may also calculate $\ddot{\sigma}^2$ as [15]: $\ddot{\sigma}^2 = \int \omega^4 S(\omega) d\omega = \frac{\pi^4}{5} W^4\sigma^2$.

By using (4.1) and (4.72), we have

$$E\{x_R\ddot{x}_I\} = E\{x_I\ddot{x}_R\} = 0,$$
$$E\{x_R\ddot{x}_R\} = E\{x_I\ddot{x}_I\} = -E\{\dot{x}_I^2\} = -\frac{\pi^2}{3}W^2\sigma^2,$$

and

$$\rho_{x_R\ddot{x}_R} = \rho_{x_I\ddot{x}_I} = -\frac{\sqrt{5}}{3},$$

Thus, given x_R and x_I, $\ddot{x}_R$ and $\ddot{x}_I$ are independent Gaussian processes with mean $-\frac{\pi^2}{3}W^2x_R$ (for $\ddot{x}_R$) and $-\frac{\pi^2}{3}W^2x_I$ (for $\ddot{x}_I$) and variances of $\frac{4}{9}\ddot{\sigma}^2$.

Note that $\dot{x}_I = \gamma|x_R|$. Given $\dot{r} = \frac{1}{r}(x_R\dot{x}_R + x_I\dot{x}_I) = 0$, we have $\dot{x}_R = -\frac{x_I\dot{x}_I}{x_R}$, and

$$\ddot{r} = \frac{1}{r}(\dot{x}_R^2 + \dot{x}_I^2 + x_R\ddot{x}_R + x_I\ddot{x}_I) = \frac{1}{r}(r^2\gamma^2 + x_R\ddot{x}_R + x_I\ddot{x}_I). \tag{4.73}$$

Then, given x_R, x_I, γ, and $\dot{r} = 0$, $\ddot{r}$ is a Gaussian process with mean

$$\hat{m}_{\ddot{r}} = E\{\ddot{r}|x_R, x_I, \gamma, \dot{r} = 0\} = \frac{1}{r}\left(r^2\gamma^2 - \frac{\pi^2}{3}W^2(x_R^2 + x_I^2)\right) = r(\gamma^2 - \frac{\pi^2W^2}{3}),$$

and variance

$$\begin{aligned}\sigma_{\ddot{r}}^2 &= E\{\ddot{r}^2|x_R, x_I, \gamma, \dot{r} = 0\} - \hat{m}_{\ddot{r}}^2 \\ &= \frac{4}{9}\ddot{\sigma}^2 = \frac{4\pi^4}{45}W^4\sigma^2.\end{aligned}$$

Because $\sigma_{\ddot{r}}^2$ is independent of x_R, x_I and γ, it is also the variance of $\ddot{r}$ given $\dot{r} = 0$ and $r \geq A$. By normalizing $\ddot{r}$ by dividing it by σ, the variance of $\ddot{r}/\sigma$ given $\dot{r} = 0$ and $r \geq A$ is $\sigma_{\ddot{r}}^2/\sigma^2 = \frac{4\pi^4W^4}{45}$, which is a constant related to only the OFDM bandwidth W.

$\hat{m}_{\ddot{r}}$ is also the mean of $\ddot{r}$ given r, γ and $\dot{r} = 0$ because $\hat{m}_{\ddot{r}}$ depends on r instead of the individual x_R and x_I. Then, we have

$$\begin{aligned}p(\ddot{r}|\gamma, r, \dot{r} = 0) &= p(\ddot{r}|\gamma, x_R, x_I, \dot{r} = 0) \\ &= \frac{1}{\sqrt{2\pi}\sigma_{\ddot{r}}}\exp\left(-\frac{(\ddot{r} - \hat{m}_{\ddot{r}})^2}{2\sigma_{\ddot{r}}^2}\right).\end{aligned} \tag{4.74}$$

The conditional mean of $\ddot{r}/\sigma$ given $\dot{r} = 0$ and $r \geq A$ can be calculated as

$$\begin{aligned}m_{\ddot{r}/\sigma} &= E\{\frac{\ddot{r}}{\sigma}|\dot{r} = 0, r \geq A\} \\ &= \frac{\int_{-\infty}^{\infty}\int_{A}^{\infty}\int_{-\infty}^{\infty}\int_{0}^{2\pi}\ddot{r}p(\ddot{r}, \gamma, r, \theta|\dot{r} = 0)d\theta d\gamma dr d\ddot{r}}{\sigma\int_{A}^{\infty}p(r)dr}.\end{aligned}$$

However,

$$\int_{-\infty}^{\infty} \ddot{r} p(\ddot{r}, \gamma, r, \theta | \dot{r} = 0) d\ddot{r} = \int_{-\infty}^{\infty} \ddot{r} p(\ddot{r} | \gamma, r, \theta, \dot{r} = 0) p(\gamma, r, \theta | \dot{r} = 0) d\ddot{r}$$
$$= \hat{m}_{\ddot{r}} p(\gamma, r, \theta | \dot{r} = 0),$$

where $p(\gamma, r, \theta | \dot{r} = 0)$ is calculated in (4.67). Therefore,

$$m_{\ddot{r}/\sigma} = \frac{\int_A^{\infty} \int_{-\infty}^{\infty} \int_0^{2\pi} \hat{m}_{\ddot{r}} p(\gamma, r, \theta | \dot{r} = 0) d\theta d\gamma dr}{\sigma \int_A^{\infty} p(r) dr}$$
$$= \frac{\sqrt{2\pi}(\sqrt{5}\sigma\ddot{\sigma} - 3\dot{\sigma}^2)\text{erfc}\left(\frac{A}{\sqrt{2}\sigma}\right) e^{A^2/2\sigma^2} - 2\sqrt{5}\ddot{\sigma}A}{6\sigma^2}$$
$$= -\frac{\pi^2 W^2 A}{3\sigma}.$$

When $A \to \infty$, $m_{\ddot{r}/\sigma} \to -\infty$. Because $\ddot{r}/\sigma$ has a constant (conditional) variance for any A, we have $\Pr[\ddot{r} > 0 | \dot{r} = 0, r \geq A] \to 0$ when $A \to \infty$. Moreover, the OFDM bandwidth W is usually large (several MHz). Thus, unless A is very small, $m_{\ddot{r}/\sigma} \ll 0$, implying that $\Pr[\ddot{r} > 0 | \dot{r} = 0, r \geq A] \approx 0$.

4.7.D Given $\dot{r} = 0$ and $r \geq A$, γ_i and τ_i Are Uncorrelated

We first find the conditional joint pdf $p(\ddot{r}, \gamma, r | \dot{r} = 0)$. By using (4.67) and (4.74), we have

$$p(\ddot{r}, \gamma, r | \dot{r} = 0) = p(\ddot{r} | \gamma, r, \theta, \dot{r} = 0) p(\gamma, r, \theta | \dot{r} = 0) p(\theta)$$
$$= \frac{r^2}{2\pi\sigma_{\ddot{r}}\dot{\sigma}\sigma^2} \exp\left(-\frac{(\ddot{r} - m_{\ddot{r}})^2}{2\sigma_{\ddot{r}}^2} - \frac{r^2\gamma^2}{2\dot{\sigma}^2} - \frac{r^2}{2\sigma^2}\right).$$

Because $\tau = \sqrt{-\frac{8(r - A)}{\ddot{r}}}$, the conditional joint moments of τ and γ can be found as

$$E\{\tau^m \gamma^n | \dot{r} = 0, r \geq A\} = \frac{\int_A^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{0} \tau^m \gamma^n p(\ddot{r}, \gamma, r | \dot{r} = 0) d\ddot{r} d\gamma dr}{\int_A^{\infty} p(r) dr},$$

where m and n are positive integers. However, note that $p(\ddot{r}, \gamma, r | \dot{r} = 0)$ is symmetric to γ. Therefore,

$$\int_{-\infty}^{\infty} \tau^m \gamma^n p(\ddot{r}, \gamma, r | \dot{r} = 0) d\gamma = 0$$

for any odd n, and, in turn, $E\{\tau^m \gamma^n | \dot{r} = 0, r \geq A\} = 0$ for any odd n. Specifically, $E\{\tau\gamma | \dot{r} = 0, r \geq A\} = 0$. That is, given $\dot{r} = 0$ and $r \geq A$, τ and γ are uncorrelated.

4.7.E The Conditional pdf and moments of η_k

In this section, we find the conditional mean and variance of

$$\eta_k = \frac{\dot{x}_I(t_k)r(t_k) - x_I(t_k)\dot{r}(t_k)}{r(t_k)|x_R(t_k)|},$$

given $\dot{r}(t_k) \geq 0$ and $r(t_k) = A$. In the following, we drop off the time index t_k for the ease of notation.

Note that

$$p(x_R, x_I, \dot{x}_I, \dot{r}) = p(\dot{x}_I, \dot{r}|x_R, x_I)p(x_R)p(\dot{x}_I),$$

where $p(\dot{x}_I, \dot{r}|x_R, x_I)$ is defined in (4.61). We have

$$\begin{aligned} p(\dot{x}_I, \dot{r}, \theta|r) &= \frac{rp(x_R = r\cos\theta, x_I = r\cos\theta, \dot{x}_I, \dot{r})}{p(r)} \\ &= \frac{1}{4\pi^2\dot{\sigma}^2|\cos\theta|}\exp\left(-\frac{\dot{x}_I^2 - 2\dot{x}_I\dot{r}\sin\theta + \dot{r}^2}{2\dot{\sigma}^2\cos^2\theta}\right). \end{aligned}$$

The Jacobian of transforming $p(\dot{x}_I, \dot{r}, \theta|r)$ to $p(\eta_k, \dot{r}, \theta|r)$ is $r|\cos\theta|$ because $\dot{x}_I = r\eta_k|\cos\theta| + \dot{r}\sin\theta$. Then, the pdf of η_k conditioned on $\dot{r} \geq 0$ and $r = A$ is

$$\begin{aligned} p(\eta_k|\dot{r} \geq 0, r = A) &= \frac{\int_0^\infty \int_0^{2\pi} r|\cos\theta|p(\dot{x}_I = r\eta_k|\cos\theta| + \dot{r}\sin\theta, \dot{r}, \theta|r)d\theta d\dot{r}}{\int_0^\infty p(\dot{r})d\dot{r}} \\ &= \frac{\sqrt{6}A}{2\pi\sqrt{\pi}W\sigma}\exp\left(-\frac{3A^2\eta_k^2}{2\pi^2W^2\sigma^2}\right), \end{aligned}$$

The conditional moments of η_k can be found:

$$\begin{aligned} m_{\eta_k} &= E\{\eta_k|\dot{r} \geq 0, r = A\} = 0, \\ \sigma_{\eta_k}^2 &= E\{\eta_k^2|\dot{r} \geq 0, r = A\} = \frac{\dot{\sigma}^2}{A^2} = \frac{\pi^2W^2\sigma^2}{3A^2}, \\ \sigma_{\eta_k}^4 &= E\{\eta_k^4|\dot{r} \geq 0, r = A\} = \frac{\pi^4W^4\sigma^4}{3A^4}, \\ m_{|\eta_k|} &= E\left\{|\eta_k||\dot{r} \geq 0, r = A\right\} = \sqrt{\frac{2\pi}{3}}\frac{W\sigma}{A}, \end{aligned}$$

and the conditional variance of $|\eta_k|$ is

$$\sigma_{|\eta_k|}^2 = \sigma_{\eta_k}^2 - m_{|\eta_k|}^2 = \left(\frac{\pi^2}{3} - \frac{2\pi}{3}\right)\frac{W^2\sigma^2}{A^2}. \tag{4.75}$$

When A is large, η_k has approximately the same distribution as γ_i (calculated in Appendix 4.7.B). (4.68) shows that $\psi \gg 1$ when A is large. Then,

$$\sqrt{\pi}\text{erfc}\,(\psi) \approx \frac{e^{-\psi^2}}{\psi} \ll 2\psi e^{-\psi^2}.$$

Thus, (4.68) can be approximated as

$$p(\gamma_i|\dot{r}=0, r\geq A) \approx \frac{A^3\psi e^{-\psi^2} e^{A^2/2\sigma^2}}{2\sqrt{2\pi}\dot{\sigma}\sigma^2\psi^3}$$
$$= \frac{\pi^2W^2}{3\gamma_i^2+\pi^2W^2}\frac{\sqrt{6}A}{2\pi\sqrt{\pi}W\sigma}\exp\left(-\frac{3A^2\gamma_i^2}{2\pi^2W^2\sigma^2}\right)$$
$$\approx \frac{\sqrt{6}A}{2\pi\sqrt{\pi}W\sigma}\exp\left(-\frac{3A^2\gamma_i^2}{2\pi^2W^2\sigma^2}\right),$$

which is the same as $p(\eta_k|\dot{r}\geq 0, r=A)$. This approximation is valid when $3\gamma_i^2 \ll \pi^2W^2$. On the other hand, when $3\gamma_i^2$ is comparable to or larger than π^2W^2, both $p(\gamma_i|\dot{r}=0, r\geq A)$ and this approximation are virtually zero. Therefore, we may use this approximation for all γ_i.

Remark 4.8. By using the asymptotic expansion of $\mathrm{erfc}\,(x)$ and $\mathrm{E}_1\,(x)$ [211], it is easy to verify that the moments of η_k are approximately the same as those of γ_i when A is large. Intuitively, we can explain this fact by observing that η_k is measured at the time that $r(t)$ up-crosses the level A, and γ_i is measured at the time that $r(t)$ reaches its local peak after the up-crossing. Because these two time instances are very close when A is large, the statistics of η_k and γ_i are the same.

4.8 The Execution Time of DIT Split-Radix JN-Point FFT with N_f Nonzero Inputs and N Desired Outputs

In this section, we analyze the execution time of calculating the JN-point FFT with N_f nonzero inputs and N desired outputs by using the DIT algorithm. The execution time is measured as the number of real multiplications. A complex multiplication is equivalent to three real multiplications [197], except one of the multipliers is 1 (no multiplication needed) or $(1+j)\frac{\sqrt{2}}{2}$ (two real multiplications).

Consider the DFT

$$X_k = \sum_{n=0}^{JN-1} x_n e^{-j2\pi nk/JN},$$

where $n = 0, ..., JN-1$, and $k = 0, ..., N/2-1, JN-N/2, ..., JN-1$ (which are the desired outputs of the DFT corresponding to the N inband frequency-domain samples of the N-subcarrier OFDM system). The input x_n has only N_f nonzero

values randomly distributed on $n = 0, ..., JN-1$. We assume that both J and N are the power of 2.

Define $W_{JN} = e^{j2\pi/JN}$. By using the DIT decomposition, the split-radix FFT can be written as [197]

$$X_{2k} = \sum_{n=0}^{JN/2-1} (x_n + x_{n+JN/2}) W_{JN}^{2nk} \tag{4.76}$$

$$X_{4k+1} = \sum_{n=0}^{JN/4-1} ((x_n - x_{n+JN/2}) - j(x_{n+JN/4} - x_{n+3JN/4})) W_{JN}^{n} W_{JN}^{4nk} \tag{4.77}$$

$$X_{4k+1} = \sum_{n=0}^{JN/4-1} ((x_n - x_{n+JN/2}) + j(x_{n+JN/4} - x_{n+3JN/4})) W_{JN}^{3n} W_{JN}^{4nk}. \tag{4.78}$$

Eq. (4.76) is a $(JN/2)$-point DFT with $N/2$ desired outputs ($k = 0, ..., \frac{N}{4} - 1, \frac{JN}{2} - \frac{N}{4}, ..., \frac{JN}{2} - 1$). The minimum number of nonzero inputs is $\min(N_f/2, JN/2)$, and the maximum number of nonzero inputs is $\min(N_f, JN/2)$. Here, we assume the x_n's are distributed in a way such that all the DFT involved in the DIT decomposition have the maximum number of nonzero inputs, leading to the worst (i.e., the largest) execution time.

Eqs. (4.77) and (4.78) are two $(JN/4)$-point DFT with $N/4$ desired outputs ($k = 0, ..., N/8 - 1, JN/4 - N/8, ..., JN/4 - 1$), and $\min(N_f, JN/4)$ nonzero inputs. Calculating each input requires three real multiplications except for $n = 0$ (no multiplication is needed) and $n = JN/8$ (which requires two real multiplications). However, the 0-th and $(JN/8)$-th inputs may be zero unless $N_f \geq JN/4$. Therefore, in view of the worst execution time, calculating (4.76) – (4.78) requires

$$\mathcal{M}_{JN} = \mathcal{M}_{JN/2} + 2\mathcal{M}_{JN/4} + \max(0, \min(6N_f, 3JN/2 - 8)) \tag{4.79}$$

real multiplications, where $\mathcal{M}_{JN}$, $\mathcal{M}_{JN/2}$ and $\mathcal{M}_{JN/4}$ are the number of real multiplications for calculating the (JN)-, $(JN/2)$- and $(JN/4)$-point DFT, respectively.

The DIT decomposition continues until each $(2J)$-point DFT decomposes to one J-point DFT and two $(J/2)$-point DFT. The J-point DFT has only one desired output at $k = 0$. Therefore, its calculation needs no multiplications, i.e.,

$$\mathcal{M}_J = 0. \tag{4.80}$$

The $(J/2)$-point DFT corresponding to (4.77) has no desired outputs and can be omitted. On the other hand, the $(J/2)$-point DFT corresponding to (4.78) has one desired output at $k = J/2 - 1$, and $\min(N_f, J/2)$ nonzero inputs. By pre-computing and storing W_{2J}^{n+4nk} and W_{2J}^{3n+4nk} ($n = 1, ..., J/2-1$ and $k = J/2-1$), (4.78) requires $\max(0, \min(3N_f, 3J/2 - 4))$ real multiplications. Therefore,

$$\mathcal{M}_{2J} = \max(0, \min(3N_f, 3J/2 - 4)). \tag{4.81}$$

Now, the execution time of calculating the JN-point FFT with N_f nonzero inputs and N desired outputs by using DIT can be found by recursively using (4.79) and applying the initial conditions (4.80) and (4.81).

From the above analysis, we see that the overall execution time is a function of N_f. If N_f is a random variable, we may estimate the averaged FFT execution time by replacing N_f in (4.79) and (4.81) by its mean $\bar{N}_f$.

Chapter 5

PAR Reduction with Discrete Solutions

This chapter focuses on discrete PAR-reduction techniques, which modify the phase and/or amplitude of the data symbols in a discrete manner to reduce the PAR. We first propose a method to eliminate the side information for the sign-selection technique. This method optimizes the signs of the data symbols to minimize the PAR. Then, we propose new probabilistic algorithms to find better suboptimal solutions for discrete PAR-reduction techniques. In this chapter, we use the zero-padding OFDM system where the carrier frequency is on the left end of the OFDM frequency band. The time-domain OFDM symbol $x(t)$ and its discrete-time samples x_n may be written as

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j2\pi kt/T}, \quad 0 \le t \le T, \tag{5.1}$$

where N data symbols X_k form an OFDM block $\mathbf{X} = [X_0, ..., X_{N-1}]$, T is the OFDM symbol period, and

$$x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k e^{j\frac{2\pi}{JN}nk}, \quad n = 0, ..., JN-1, \tag{5.2}$$

where J is the oversampling factor.

5.1 Characterization of Discrete Solutions

Discrete PAR-reduction techniques use a set of candidates to represent an OFDM block, and choose the one leading to the lowest PAR for transmission. In other

words, given an input OFDM block $\mathbf{X}$, we form a candidate set $\mathcal{X}_k$ for each data symbol X_k, $k = 0, ..., N-1$; i.e.,

$$\mathcal{X}_k = \{X_{k,0}, X_{k,1}, ..., X_{k,I_k-1}\}, \quad k = 0, ..., N-1, \tag{5.3}$$

where I_k is the size of $\mathcal{X}_k$, and $X_{k,i}$ is determined by X_k and the predefined rules of the candidate set. For example, the multiple representation technique uses

$$X_{k,i} = X_k e^{j\phi_i}, \quad i = 0, ..., I-1,$$

where ϕ_i's are a set of predefined phases. The data symbols $\hat{X}_k$ of the transmitted OFDM block are chosen from $\mathcal{X}_k$ such that the PAR is minimized. Thus, the objective function of discrete PAR-reduction techniques can be written as

$$\min_{\hat{X}_0,...,\hat{X}_{N-1}} \frac{\max |x_n|^2}{\mathrm{E}\left\{|x_n|^2\right\}}, \tag{5.4a}$$

$$\text{where } x_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{X}_k e^{\frac{j2\pi nk}{JN}}, \quad n = 0, ..., N-1, \tag{5.4b}$$

$$\text{subject to } \hat{X}_k \in \mathcal{X}_k, \quad i = 0, ..., N-1. \tag{5.4c}$$

The denominator in (5.4a) may be omitted if the average power $\mathrm{E}\left\{|x_n|^2\right\}$ does not change with the choice of X_k. The search space of this optimization can be written as

$$\Lambda = \prod_{k=0}^{N-1} I_k.$$

For example, the sign-selection technique optimizes the signs of the data symbols to minimize the PAR. That is,

$$\mathcal{X}_k = \{X_k, -X_k\}. \tag{5.5}$$

The PAR of an OFDM block $\mathbf{X} = \{X_0, ..., X_{N-1}\}$ is equal to the PAR of $-\mathbf{X}$. Therefore, we may always fix the sign of one data symbol, e.g., X_0, to be 1. Thus, $\mathcal{X}_0 = \{X_0\}$, $I_0 = 1$, and $I_k = 2$ for $k \neq 0$, and the search space is $\Lambda = 2^{N-1}$.

If the elements of $\mathcal{X}_k$ belong to the same constellation, the index of the selected candidate must be transmitted to the receiver as side information in order to correctly decode the data symbols. The uncoded side information requires

$$B_s = \sum_{k=0}^{N-1} \log_2 I_k \text{ (bits)}.$$

For example, the side information of the sign-selection technique is $B_s = N - 1$ bits.

5.2 PAR Reduction By Using Adaptive Mapping

The adaptive mapping technique for PAR reduction is based on the fact that, for an arbitrarily given signal constellation, some data sequences have a low PAR [1]. By using the adaptive mapping technique, an M-point signal constellation $\mathcal{Q}_M$ can be mapped to a $2M$-point signal constellation $\mathcal{Q}_{2M}$. In this case, each point in $\mathcal{Q}_M$ can be represented by either of the two corresponding points in $\mathcal{Q}_{2M}$. That is, each OFDM subcarrier has two modulation choices. The extra freedom provided by this method allows each subcarrier modulation symbol to be carefully selected to reduce the PAR. On the other hand, in order to decode the received signal easily and reliably (such decoding is especially important for the mobile handset because of its low cost requirement), the mapping scheme has to be simple and revertible without any side information.

Given the above consideration, we propose an adaptive mapping technique that involves only sign-selection to map $\mathcal{Q}_M$ to $\mathcal{Q}_{2M}$.

Let q_m, $m = 1, ..., 2M$, represent an arbitrary point in $\mathcal{Q}_{2M}$. For any origin-symmetrical signal constellation such as PSK or QAM,

$$q_{\langle m+M \rangle_{2M}} = -q_m \in \mathcal{Q}_{2M}, \tag{5.6}$$

where $\langle \cdot \rangle_{2M}$ represents modulo-$2M$ operation. Then, $\mathcal{Q}_{2M}$ is divided into two disjointed sets by choosing M adjacent points $\{q_m, m = 1, ..., M\}$ to form a subset $\mathcal{Q}_{2M}^{(1)}$, and choosing the remaining M adjacent points $\{q_{m+M}, m = 1, ..., M\} = \{-q_m, m = 1, ..., M\}$ to form another subset $\mathcal{Q}_{2M}^{(2)}$.

Now, given an arbitrary point $\hat{q}_m \in \mathcal{Q}_M$, we can map it to either $q_m \in \mathcal{Q}_{2M}^{(1)}$ or $-q_m \in \mathcal{Q}_{2M}^{(2)}$. As an example, the mapping from QPSK to 8PSK is shown in Fig. 5.1, where 1 in QPSK is mapped to either -1 or 1 in 8PSK, and -1 in QPSK is mapped to either $-j$ or j in QPSK, and so on.

For any $q_m \in \mathcal{Q}_{2M}^{(1)}$ $(m = 1, ..., M)$, $-q_m$ is not a member of $\mathcal{Q}_{2M}^{(1)}$. Therefore, the inverse mapping from $\mathcal{Q}_{2M}^{(1)}$ (or $\mathcal{Q}_{2M}^{(2)}$) to $\mathcal{Q}_M$ is unique, and no side information is needed for decoding.

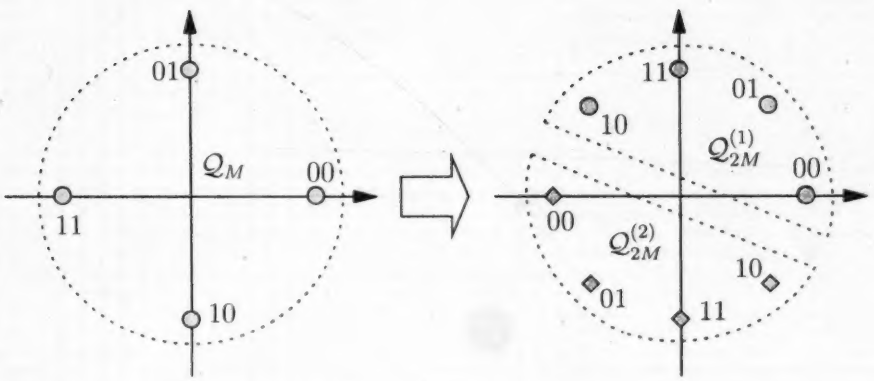


Figure 5.1: QPSK mapping to 8PSK

With this mapping scheme, we first encode each $\log_2 M$ input bits to a point in, say, $\mathcal{Q}_{2M}^{(1)}$ to obtain $\mathbf{X} = [X_0, ..., X_{N-1}]$. We associate each X_k with a sign-selection variable s_k. The original OFDM block $\mathbf{X}$ is thus replaced by

$$\hat{\mathbf{X}} = [s_0 X_0, ..., s_{N-1} X_{N-1}], \tag{5.7}$$

and the discrete-time transmit signal is given by

$$\hat{x}_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k X_k e^{j2\pi \frac{nk}{JN}}, \quad n = 0, ..., JN-1. \tag{5.8}$$

The choice of s_k does not affect the average power of $\hat{x}_n$, which is the same as that of x_n. The PAR-reduction problem is therefore equivalent to minimizing the maximum peak of the amplitude of $\hat{x}_n$. Consequently, this problem is reformulated as

$$\min_{\mathbf{s}} \max_{n=0,...,JN-1} \left| \sum_{k=0}^{N-1} s_k X_k e^{j2\pi \frac{nk}{JN}} \right|^2 \tag{5.9}$$
$$\text{subject to: } \mathbf{s} \in \{1, -1\}^N,$$

where $\mathbf{s} = [s_0, ..., s_{N-1}]$, and $\{1, -1\}^N$ is the set of N-dimensional binary vectors. For any sign sequence $\mathbf{s}$, a sign sequence $-\mathbf{s}$ always exists that leads to the same PAR. Therefore, we can always let $s_0 \equiv 1$. The redundancy introduced by this technique is $1/\log_2(2M) = 1/(1 + \log_2 M)$. By using a large constellation, we can keep the redundancy small (e.g., 16.67% for $M = 32$).

The adaptive mapping technique proposed above can significantly reduce the PAR. For example, the PAR distribution for a 16-subcarrier OFDM system ($N = 16$) with a

Binary Phase Shift Keying (BPSK) input ($M = 2$) is investigated. The total number of possible input OFDM blocks is $2^{16} = 65536$. After mapping BPSK to QPSK by using the adaptive mapping technique, the search space of each input OFDM block is $2^{15} = 32768$. Fig. 5.2 shows the PAR distribution of ordinary BPSK OFDM blocks and that of the optimum adaptively-mapped QPSK OFDM blocks found by using the exhaustive search. For ordinary BPSK OFDM blocks, the PAR is 11 dB at a clip probability of 10^{-4}, and the worst PAR is $10 \log_{10} N = 12.04$ dB. However, after adaptive mapping, the optimized QPSK OFDM blocks has the worst PAR of only 3.5 dB (with a clip probability of 1.22×10^{-4}). Therefore, the adaptive mapping technique can reduce the PAR significantly.

Finding the optimal sign sequence for each OFDM block requires an exhaustive search among all 2^{N-1} possible choices, where testing each choice requires an IFFT operation. Thus, the complexity of the exhaustive search is a exponential function of the number of subcarriers, making an exhaustive search impractical when N is large. For example, $2^{127} \approx 1.7 \times 10^{38}$ IFFTs are required when $N = 128$. Many algorithms have been proposed in the literature to find the suboptimal solutions of the discrete optimization problem (5.9). Generally, the execution time of finding a nearly optimal solution of (5.9) may be prohibitively high. For example, to limit the PAR of a 128-subcarrier 16QAM-modulated OFDM system to no larger than 6 dB at a clip probability of 10^{-4}, SLM requires an average of 556 IFFTs, and 7425 IFFTs in the worst case of our simulation, where 10^5 OFDM blocks are tested.

The existing algorithms can be divided into two groups: (1) Algorithms with low-complexity , e.g., SLM and PTS [13,113], which, however, obtain only moderate PAR reduction; and (2) algorithms with near-optimal solutions, e.g., the derandomization method [22], which have relatively high complexity ($\mathcal{O}(N^2)$) but may serve as the foundation of future fast algorithms.

In the rest of this chapter, we propose several suboptimal algorithms to solve (5.9). These algorithms belong to the second group for solving (5.9), i.e., for finding a near-optimal solution. These algorithms also offer a flexible tradeoff between PAR reduction and execution time. Our simulations show that, for the same level of PAR reduction, our algorithms require less execution time than the SLM and derandomization methods. In the last section of this chapter, two fast algorithms are proposed,

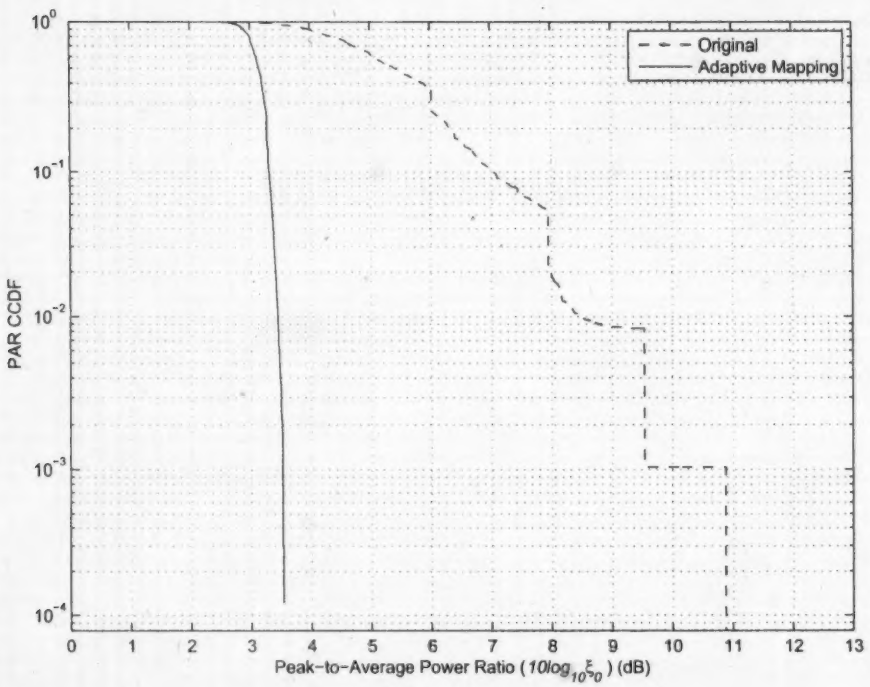


Figure 5.2: PAR for uncoded BPSK symbol sequences and for adaptive-mapping coded symbol sequences, where $N = 16$, $M = 2$.

which reduce the PAR of a 128-subcarrier OFDM system to about 6 dB at a clip probability of 10^{-4} with only 16 IFFTs.

5.3 Probabilistic Solutions

5.3.1 Random Selection Method

The first method is based on the random selection of $\mathbf{s}$. We first randomly select K sign sequences $\mathbf{s}^{(1)}$, $\mathbf{s}^{(2)}$, ..., $\mathbf{s}^{(K)}$. For each OFDM block $\mathbf{X}$, the maximum instantaneous power (or PAR) of each $\mathbf{s}^{(l)} \odot \mathbf{X} = [s_0^{(l)} X_0, ..., s_{N-1}^{(l)} X_{N-1}]$, where $l = 1, ..., K$ and $\odot$ represents element-wise multiplication, is calculated. The one with the lowest maximum instantaneous power is then selected for transmission. The search space is 2^K. A tradeoff between K and the PAR requirement can be made.

In (5.9), $\mathbf{X}$ comes from half of the $(2M)$-ary constellation. By using the random selected sign sequence $\mathbf{s}$, the adaptively mapped OFDM block becomes a sequence coming from the $(2M)$-ary constellation $\mathcal{Q}_{2M}$. Therefore, this method is equivalent to

randomly selecting a sequence from $\mathcal{Q}_{2M}$. Because the PAR distribution of $\mathcal{Q}_{2M}$ symbol sequences is exponentially distributed, this method can achieve only a moderate PAR reduction for a moderate K.

5.3.2 Modified PTS Algorithms

The PTS technique [13] partitions the input sequence into several subgroups, called partial transmit sequences. A sign sequence, where each element corresponds to the sign for each subgroup, is used to obtain a suboptimal solution. Several algorithms for the PTS technique have been proposed in the literature [28]. With some modifications, these algorithms can also be used here.

As mentioned above, $\mathbf{X}$ comes from $\mathcal{Q}_{2M}^{(1)}$, which is a non-origin-symmetrical constellation. This fact implies that its PAR distribution is much higher than that of $\mathcal{Q}_{2M}$, an origin-symmetrical constellation (see Fig. 5.3). Therefore, directly applying a PTS algorithm to $\mathbf{X}$ would not achieve a significant PAR reduction compared to applying it to the original M-ary constellation symbol sequences.

Because PTS algorithms work well for origin-symmetrical constellation symbol sequences, the modification is simple. First, randomly generate an initial sign sequence, denoted as $\mathbf{s}_{\text{ini}}$, such that $\mathbf{X}$ is mapped to a $\mathcal{Q}_{2M}$ symbol sequence $\mathbf{X}_{\text{ini}} = \mathbf{s}_{\text{ini}} \odot \mathbf{X}$. Then, PTS algorithms can be used on $\mathbf{X}_{\text{ini}}$ to reduce its PAR.

As an example, Fig. 5.3 illustrates the PAR distribution for (1) original $\mathcal{Q}_{2M}^{(1)}$ symbol sequences (– –), (2) original $\mathcal{Q}_{2M}$ symbol sequences (—), (3) $\mathcal{Q}_{2M}^{(1)}$ symbol sequences with PTS (– ·), and (4) $\mathcal{Q}_{2M}$ symbol sequences with PTS ($\cdots$). Here, an OFDM system of $N = 64$ subcarriers and 8PSK ($M = 8$) data input is used. When PTS is used, each OFDM block is partitioned into four subgroups. The PAR for $\mathcal{Q}_{2M}^{(1)}$ symbol sequences is quite high no matter if the PTS technique is used or not. On the other hand, with the use of the initial sign sequence $\mathbf{s}_{\text{ini}}$, the PAR can be significantly lowered.

5.3.3 Recursive Partial Sequence Method

This method is based on the recursive optimization of the partial symbol sequence. Given a OFDM block $\mathbf{X}$ (from $\mathcal{Q}_{2M}^{(1)}$), one can partition it into L_g sub-blocks with

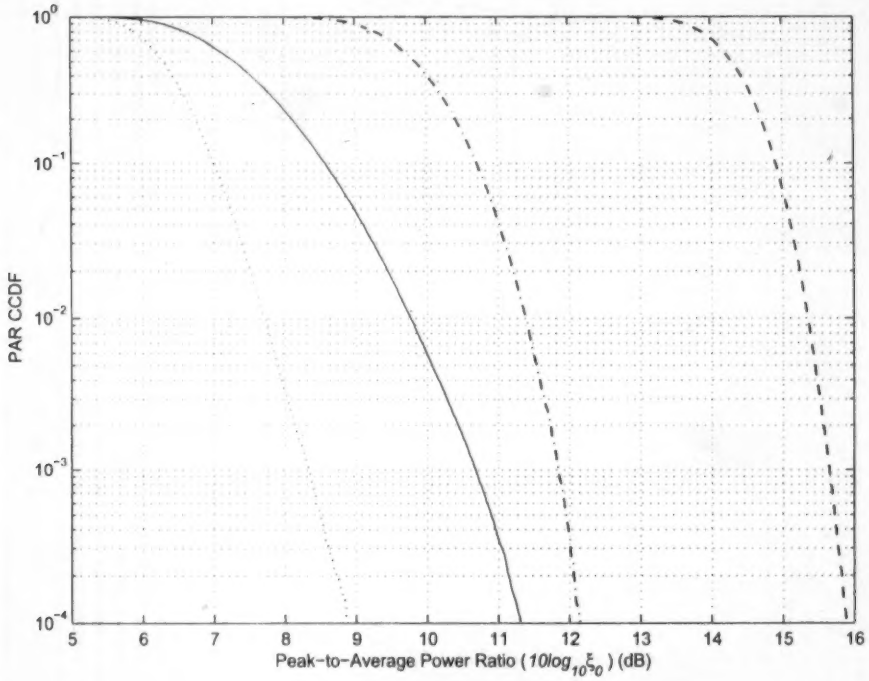


Figure 5.3: PAR CCDF for (1) original $\mathcal{Q}_{2M}^{(1)}$ symbol sequences (– –), (2) original $\mathcal{Q}_{2M}$ symbol sequences (—), (3) $\mathcal{Q}_{2M}^{(1)}$ symbol sequences with PTS (– ·), and (4) $\mathcal{Q}_{2M}$ symbol sequences with PTS (· · ·), where $N = 64$, $M = 8$, and each OFDM block is partitioned into four subgroups when using PTS.

each sub-block containing $K = N/L_g$ symbols:

$$\mathbf{X} = [\mathbf{X}_0, ..., \mathbf{X}_{L_g-1}]. \tag{5.10}$$

Similarly, $\mathbf{s}$ can also be partitioned into L_g sub-blocks:

$$\mathbf{s} = [s_0, ..., s_{N-1}] = [\mathbf{s}_0, ..., \mathbf{s}_{L_g-1}]. \tag{5.11}$$

Let $\mathcal{S}_k$ denote the symbol index set of the sub-block $\mathbf{X}_k$ or $\mathbf{s}_k$. Then $\mathcal{S}_k$ satisfies the following relationships:

$$\mathcal{S}_k \cap \mathcal{S}_l = \oslash, \quad k \neq l,\ k, l = 0, ..., L_g - 1, \tag{5.12}$$

$$\bigcup_{k=1}^{L_g} \mathcal{S}_k = \{1, ..., N-1\}, \tag{5.13}$$

where $\oslash$ represents the empty set. Similar to the partitioning with the PTS technique, adjacent, interleaved, and pseudo random sub-block partitioning can be used. Here,

we use adjacent partitioning. Therefore,

$$\mathcal{S}_k = \left\{ \frac{kN}{L_g}, \frac{kN}{L_g} + 1, ..., \frac{(k+1)N}{L_g} - 1 \right\}, \quad k = 0, ..., L_g - 1. \tag{5.14}$$

Now, let us form a partial symbol sequence $\tilde{\mathbf{X}} = \mathbf{X}_0$ and a corresponding sign sequence $\tilde{\mathbf{s}} = \mathbf{s}_0$. Then, the optimal $\mathbf{s}_0$, denoted as $\mathbf{s}_0^{(\text{opt})}$, can be found by minimizing the PAR of $\tilde{\mathbf{s}} \odot \tilde{\mathbf{X}}$. Then, $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{s}}$ are updated such that $\tilde{\mathbf{X}} = [\mathbf{X}_0, \mathbf{X}_1]$, and $\tilde{\mathbf{s}} = [\mathbf{s}_0^{(\text{opt})}, \mathbf{s}_1]$, respectively. Similarly, $\mathbf{s}_1^{(\text{opt})}$ can be found by minimizing the PAR of the updated $\tilde{\mathbf{s}} \odot \tilde{\mathbf{X}}$. This procedure is repeated until $\mathbf{s}_{L_g-1}^{(\text{opt})}$ is found. Generally, this method can be written as [212]

$$\begin{aligned} &\text{FOR} \quad k = 0 \quad \text{TO} \quad L_g - 1 \\ &\qquad \tilde{\mathbf{X}} = [\mathbf{X}_0, ..., \mathbf{X}_k], \quad \tilde{\mathbf{s}} = [\mathbf{s}_0, ..., \mathbf{s}_k] \\ &\qquad \mathbf{s}_k^{(\text{opt})} = \arg\min_{\mathbf{s}_k} \left(\text{PAR of } \tilde{\mathbf{s}} \odot \tilde{\mathbf{X}} \right) \\ &\text{END FOR} \end{aligned} \tag{5.15}$$

When using this method, a search for 2^K possible choices is required in each iteration. That is, 2^K (partial) IFFTs are required in each iteration except the first iteration, where 2^{K-1} IFFTs are required because $s_0 \equiv 1$. Therefore, the total number of IFFTs is

$$\text{Number of IFFTs} = \begin{cases} 2(N-1), & K = 1, \\ \frac{N}{K} 2^K - 2^{K-1}, & \text{otherwise.} \end{cases} \tag{5.16}$$

As an example, Table 5.1 lists the total numbers of IFFTs for different K. Compared to the complete search (2^{N-1} IFFTs), the proposed method significantly reduces the complexity. Moreover, because most IFFTs involved in this method are partial IFFTs (i.e., most inputs of the IFFT are zeros), the execution time can be further reduced by using fast IFFT algorithms exploiting this fact.

5.3.4 Simulation Results

In this section, numerical simulations are performed for an OFDM system with 256 subcarriers. 10^5 8PSK ($M = 8$) input OFDM blocks are generated and then are mapped to 16PSK blocks by using adaptive mapping. We compare the PAR reduction performance of the recursive partial sequence method with that of PTS. When using the recursive partial sequence method, we consider three cases where the sub-block

Table 5.1: The total numbers of IFFTs for different K

K	Number of IFFT
1	$2N-2$
2	$2N-2$
4	$4N-8$
8	$32N-128$

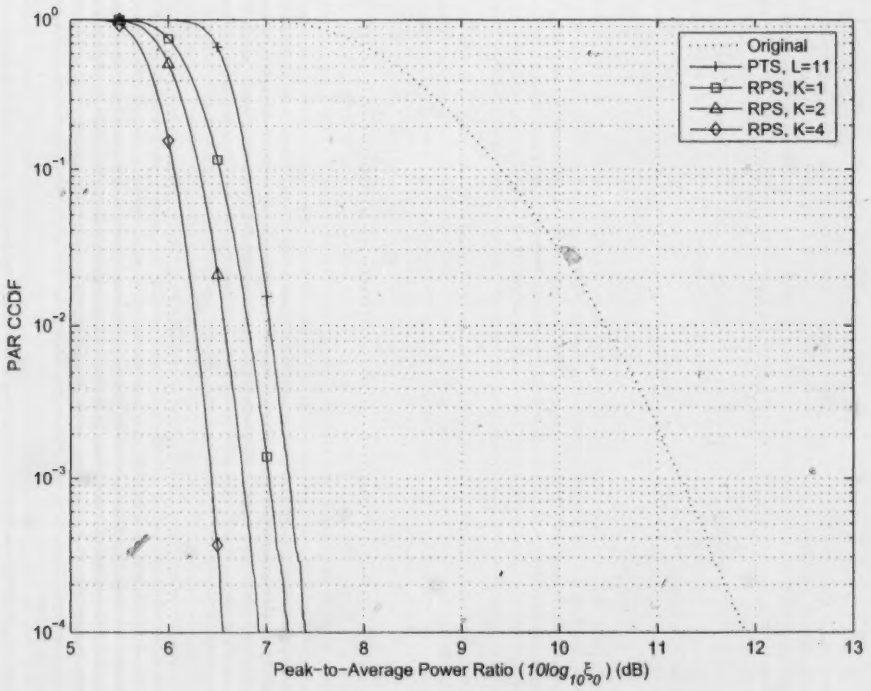


Figure 5.4: PAR comparison, where $N = 256$, $M = 8$.

length $K = 1, 2, 4$, respectively. When using PTS, the number of sub-blocks is $L = 11$.

Fig. 5.4 shows the PAR distribution for both methods. It is clear that the recursive partial sequence method can significantly reduce the PAR. When $K = 1$, the recursive partial sequence method obtains a 4.8 dB PAR reduction at a clip probability of 10^{-4}. When $K = 2$, the PAR-reduction performance is improved by 0.3 dB. From Table 5.1, the execution time of $K = 1$ and $K = 2$ is the same. Therefore, $K = 2$ is a better choice in practical situations. By increasing the execution time, the recursive partial

sequence method can obtain about a 5.5 dB PAR reduction for $K = 4$, whereas 1016 (partial) IFFTs are required to find the (sub)optimal sign sequences.

The performance of the proposed technique is better than that of PTS. In our simulations, the total number of IFFTs required for PTS is $2^{R-1} = 1024$, which is roughly the same as that for recursive partial sequence method with $K = 4$. However, the PAR reduction of PTS is about 7.4 dB at a clip probability of 10^{-4}, which is about 0.5 dB worse than that of the recursive partial sequence method with $K = 2$ and about 0.9 dB worse than that of the recursive partial sequence method with $K = 4$.

5.4 Cross-Entropy Method for PAR Reduction

Stochastic search and optimization techniques [21] can be used to find good suboptimal solutions of the discrete optimization problem (5.9), which is copied to here for the ease of reference:

$$\begin{aligned} &\min_{\mathbf{s}} \mathbb{L}(\mathbf{s}) \\ &\text{subject to: } \mathbf{s} \in \{1, -1\}^N, \end{aligned} \tag{5.17}$$

where $\mathbf{s} = [s_0, ..., s_{N-1}]$, and

$$\mathbb{L}(\mathbf{s}) = \max_{n=0,...,JN-1} \left| \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} s_k X_k e^{\frac{j2\pi nk}{JN}} \right|^2. \tag{5.18}$$

When we consider $\mathbf{s}$ as a random vector, $\mathbb{L}(\mathbf{s})$ is thus an event in the N-dimensional space $\{1, -1\}^N$. Because the occurrence of the optimal solution $\mathbf{s}^{(\text{opt})}$ has a very small probability, $\mathbb{L}(\mathbf{s}^{(\text{opt})})$ is a rare event. The random selection method described in the previous section belongs to the crude Monte-Carlo method [21], where the suboptimal $\mathbf{s}$ is found from a set of random samples $\mathbf{s}_1, ..., \mathbf{s}_n$ drawn from a (usually arbitrarily defined) distribution of $\mathbf{s}$. Other methods, e.g., simulated annealing [213–216], Tabu search [217–220], and genetic algorithms [221–224], can also be used to obtain better solutions than those obtained by using the crude Monte-Carlo method.

The CE method is an iterative procedure for combinatorial optimization [225–229]. Each iteration involves generating a random sample according to a probability distribution and then updating the parameters of the probability distribution in order to produce better samples in the next iteration [227]. In this section, we develop two CE-based PAR-reduction algorithms [230, 231]. Near-optimal solutions are obtained by simultaneously modifying the probabilities of the signs of *all* subcarriers via the CE method. In contrast, the derandomization method [22, 23] modifies the signs one-by-one. With a fixed number of iterations, our algorithms obtain the same PAR reduction with an $\mathcal{O}(N \log N)$ complexity. They also offer a flexible tradeoff between PAR reduction and execution time. The simulations show that, for the same level of PAR reduction, our algorithms requires less execution time than the SLM [13, 122] and derandomization methods. To the best of our knowledge, our work is the first application of the CE method to PAR reduction.

5.4.1 The CE Method

Let $f(\mathbf{x};\mathbf{u})$ be the probability density function (pdf) of a random vector $\mathbf{X}$ with a set of parameters $\mathbf{u}$; and $\mathbb{L}(\mathbf{X})$ represents a real-valued cost function of $\mathbf{X}$. Also, let the optimization problem be

$$\max_{\mathbf{X}} \mathbb{L}(\mathbf{X}). \tag{5.19}$$

The CE method for finding the optimal solution $\mathbf{X}^{(\text{opt})}$ involves two phases [225–227]:

1. Randomly generating a set of samples $\mathbf{X}_1, ..., \mathbf{X}_n$ with respect to a pdf $f(\mathbf{x};\mathbf{u})$, where $\mathbf{u}$ is the parameter vector to be optimized.

2. Optimizing $\mathbf{u}$ based on $\mathbb{L}(\mathbf{X}_1), ..., \mathbb{L}(\mathbf{X}_n)$ and updating $f(\mathbf{x};\mathbf{u})$ in order to produce "better" samples in the next iteration.

The optimization of $\mathbf{u}$ is to produce an optimal pdf $f(\mathbf{x};\mathbf{u}^{(\text{opt})})$ such that $\mathbf{X}^{(\text{opt})}$ occurs with large probability (this is the basic idea of importance sampling [232–235]). Fig. 5.5 shows that, if we use the Monte-Carlo method to generate a set of n samples $\mathbf{X}_1$, ..., $\mathbf{X}_n$ according to a pdf $f(\mathbf{x})$, most of these samples occur at the places where $f(\mathbf{x})$ is large. Because $\mathbf{X}^{(\text{opt})}$ is a rare event, its occurrence generally requires a large n. The ideal pdf that generates $\mathbf{X}^{(\text{opt})}$ with only one sample is

$$f^{(\text{opt})}(\mathbf{x}) = \begin{cases} 1, & \mathbf{X} = \mathbf{X}^{(\text{opt})}, \\ 0, & \text{otherwise.} \end{cases}$$

However, $f^{(\text{opt})}(\mathbf{x})$ is not feasible because $\mathbf{X}^{(\text{opt})}$ is unknown. On the other hand, if $f(\mathbf{x})$ can be controlled by a parameter $\mathbf{u}$ such that $f(\mathbf{x};\mathbf{u}_1)$ and $f(\mathbf{x};\mathbf{u}_2)$ have a similar "shape" but the different "positions" of maxima, we can "shift" $f(\mathbf{x};\mathbf{u})$ such that its maximum value is in the neighborhood of $\mathbf{X}^{(\text{opt})}$. Then, we need only a small number of samples to generate $\mathbf{X}^{(\text{opt})}$.

The CE method associates the optimization problem with a probability estimation problem. By using the Monte-Carlo method, the probability that $\mathbb{L}(\mathbf{X})$ is greater than or equal to a threshold γ may be estimated as

$$\begin{aligned} l = P_f(\mathbb{L}(\mathbf{X}) \geq \gamma) = E_f[I_{\{\mathbf{L}(\mathbf{X})\geq\gamma\}}] &= \int I_{\{\mathbf{L}(\mathbf{x})\geq\gamma\}} f(\mathbf{x};\mathbf{u}) d\mathbf{x} \\ &\approx \frac{1}{n}\sum_{i=1}^{n} I_{\{\mathbf{L}(\mathbf{X}_i)\geq\gamma\}}, \end{aligned} \tag{5.20}$$

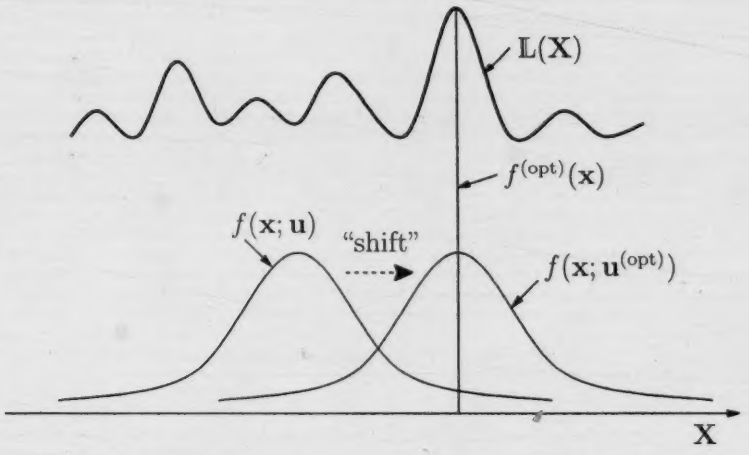


Figure 5.5: Basic idea of the CE method.

where the subscript f means that $\mathbf{X}$ is distributed with respect to the pdf f, and

$$I_{\{\mathbb{L}(\mathbf{X})\geq\gamma\}} = \begin{cases} 1, & \text{when } \mathbb{L}(\mathbf{X}) \geq \gamma, \\ 0, & \text{otherwise} \end{cases} \tag{5.21}$$

is the indicator function.

When $\mathbb{L}(\mathbf{X}_i) \geq \gamma$ is rare, estimating l requires a large n such that $\mathbb{L}(\mathbf{X}_i) \geq \gamma$ happens often. We use importance sampling to reduce n. That is, we find another pdf g for generating $\mathbf{X}$ such that $\mathbb{L}(\mathbf{X}_i) \geq \gamma$ occurs more often. By using g, (5.20) can be rewritten as

$$\begin{aligned} l &= \int I_{\{\mathbb{L}(\mathbf{x})\geq\gamma\}} \frac{f(\mathbf{x};\mathbf{u})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} = E_g[I_{\{\mathbb{L}(\mathbf{X})\geq\gamma\}} \frac{f(\mathbf{X};\mathbf{u})}{g(\mathbf{X})}] \\ &\approx \frac{1}{n} \sum_{i=1}^{n} I_{\{\mathbb{L}(\mathbf{X}_i)\geq\gamma\}} \frac{f(\mathbf{X}_i;\mathbf{u})}{g(\mathbf{X}_i)}, \end{aligned} \tag{5.22}$$

where $\mathbf{X}_i$ are generated with respect to the pdf g.

The optimal g can be obtained by minimizing the variance of l, which is

$$g^{(\mathrm{opt})}(\mathbf{x}) = \frac{I_{\{\mathbb{L}(\mathbf{X}_i)\geq\gamma\}} f(\mathbf{x};\mathbf{u})}{l}. \tag{5.23}$$

However, because (5.23) requires knowledge of l, $g^{(\mathrm{opt})}$ cannot be obtained *a priori*. The CE method approximates $g^{(\mathrm{opt})}$ with a pdf function $f(\mathbf{x};\mathbf{v})$, which belongs to the same class of $f(\mathbf{x};\mathbf{u})$ but with a different set of parameters. The optimal parameters $\mathbf{v}$ can be found by minimizing the cross entropy (or Kullback-Leibler distance) between

$g^{(\text{opt})}(\mathbf{x})$ and $f(\mathbf{x};\mathbf{v})$ as

$$\min_{\mathbf{v}} D(g^{(\text{opt})}(\mathbf{x}), f(\mathbf{x};\mathbf{v})), \tag{5.24}$$

where

$$D(g^{(\text{opt})}(\mathbf{x}), f(\mathbf{x};\mathbf{v})) = E_{g^{(\text{opt})}}\left[\ln \frac{g^{(\text{opt})}(\mathbf{x})}{f(\mathbf{x};\mathbf{v})}\right], \tag{5.25}$$

which is equivalent to

$$\begin{aligned} \max_{\mathbf{v}} D(\mathbf{v}) &= \max_{\mathbf{v}} E_{\mathbf{u}}\left[I_{\{\mathbb{L}(\mathbf{X})\geq\gamma\}} \ln f(\mathbf{X};\mathbf{v})\right] \\ &= \max_{\mathbf{v}} E_{\mathbf{w}}\left[I_{\{\mathbb{L}(\mathbf{X})\geq\gamma\}} W(\mathbf{X};\mathbf{u},\mathbf{v}) \ln f(\mathbf{X};\mathbf{v})\right], \end{aligned} \tag{5.26}$$

where

$$W(\mathbf{X};\mathbf{u},\mathbf{v}) = \frac{f(\mathbf{X};\mathbf{u})}{f(\mathbf{X};\mathbf{w})}.$$

The optimal solution $\mathbf{v}^{(\text{opt})}$ can be found by solving

$$\frac{1}{n}\sum_{i=1}^{n} I_{\{\mathbb{L}(\mathbf{X}_i)\geq\gamma\}} W(\mathbf{X}_i;\mathbf{u},\mathbf{w}) \nabla \ln f(\mathbf{X}_i;\mathbf{v}) = \mathbf{0} \tag{5.27}$$

with respect to $\mathbf{v}$. When $\mathbf{v}^{(\text{opt})}$ is found, a sample vector $\mathbf{X}^{(\text{opt})}$ can be easily generated such that $\mathbb{L}(\mathbf{X}^{(\text{opt})}) \geq \gamma$.

To ensure the efficiency of the Monte-Carlo simulation, γ has to be relatively small such that the probability of $\mathbb{L}(\mathbf{X}^{(\text{opt})}) \geq \gamma$ is not too small. On the other hand, one can predefine a relatively large probability and use it to find the corresponding γ. After finding $\mathbf{v}^{(\text{opt})}$ by using this γ, this procedure can be repeated to find a larger γ. The maximum γ and the corresponding $\mathbf{v}^{(\text{opt})}$ are then obtained when this procedure converges. At this time, the optimal $\mathbf{X}$ can be generated by using $f(\mathbf{x};\mathbf{v}^{(\text{opt})})$. The CE algorithm is summarized as follows:

1. Generate n samples $\mathbf{X}_i$ with respect to the pdf $f(\mathbf{x};\mathbf{u})$; then by using these samples, find γ such that $l \geq \rho$, where ρ is a relatively large probability (e.g., $\rho = 0.1$),

2. Let $W \equiv 1$; find an optimal $\mathbf{v}_1$ by using (5.27) and γ,

3. Replace $\mathbf{u}$ with $\mathbf{v}_1$, and repeat this procedure until some stop criteria are met.

5.4.2 The CE Sign-Selection (CESS) Algorithm for PAR Reduction

Now consider the discrete PAR-reduction problem (5.17), where the cost function $\mathbb{L}(\mathbf{s})$ that we want to *minimize* is defined in (5.18). Let $\mathbf{s} = 1 - 2\mathbf{d}$; i.e., we generate the sign sequence $\mathbf{s} \in \{+1, -1\}^N$ from a binary vector $\mathbf{d} \in \{0, 1\}^N$. Then, the cost function is

$$\mathbb{L}(\mathbf{d}) = \max_{n=0,\ldots,JN-1} \left| \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} (1 - 2d_k) X_k e^{\frac{j2\pi nk}{JN}} \right|^2. \tag{5.28}$$

Each element of $\mathbf{d}$ is modeled as an independent Bernoulli random variable with the probability distribution $P(d_k = 1) = 1 - P(d_k = 0) = p_k$, $k = 0, ..., N-1$. The probability distribution is

$$f(\mathbf{d}; \mathbf{p}) = \prod_{k=0}^{N-1} p_k^{d_k} (1 - p_k)^{1-d_k}. \tag{5.29}$$

The CE method optimizes $\mathbf{p} = [p_0, p_1, ..., p_{N-1}]$, which will generate a nearly optimal solution $\mathbf{d}^*$ that leads to a low PAR.

However, $\mathbf{d}^*$ occurs with a very small probability. Estimating this probability by using the Monte-Carlo method requires a large number of samples of $\mathbf{d}$. Instead, the CE method estimates a relatively large probability $\Pr[L(\mathbf{d}) \leq \gamma]$, where γ is a relatively large threshold. Fewer samples are required to estimate this probability. The probability vector $\mathbf{p}$ is updated so that most samples generated by $f(\mathbf{d}; \mathbf{p})$ satisfy $L(\mathbf{d}) \leq \gamma$. The likelihood of $\mathbf{d}^*$ appearing among these samples is increased. If we iteratively let $\gamma \to 0$, $f(\mathbf{d}; \mathbf{p})$ converges to an optimum pdf $f(\mathbf{d}; \mathbf{p}^*)$ that generates $\mathbf{d}^*$ with a minimum number of samples, and γ converges to $L(\mathbf{d}^*)$. In each iteration, $\mathbf{p}^*$ can be analytically found by solving [227]

$$\frac{1}{n} \sum_{i=1}^{n} I_{\{\mathbb{L}(\mathbf{d}_i) \leq \gamma\}} \nabla \ln f(\mathbf{d}_i; \mathbf{p}) = \mathbf{0}, \tag{5.30}$$

where ∇x denotes the partial derivative of x with respect to p_k,

$$I_{\{\mathbb{L}(\mathbf{d}_i) \leq \gamma\}} = \begin{cases} 1, & \mathbb{L}(\mathbf{d}_i) \leq \gamma, \\ 0, & \text{otherwise}, \end{cases}$$

and $\mathbf{d}_i$ are generated by using $f(\mathbf{d}; \mathbf{p})$.

The partial derivative of (5.29) is given by

$$\frac{\partial \ln f(\mathbf{d};\mathbf{p})}{\partial p_k} = \frac{d_k}{p_k} - \frac{1-d_k}{1-p_k}, \tag{5.31}$$

where d_k is the k-th element of $\mathbf{d}$. By substituting (5.31) into (5.30), the optimal p_k can be found as [227]

$$p_k^{(\text{opt})} = \frac{\sum_{i=1}^{n} I_{\{\mathbb{L}(\mathbf{d}_i)\leq\gamma\}} d_{ik}}{\sum_{i=1}^{n} I_{\{\mathbb{L}(\mathbf{d}_i)\leq\gamma\}}} \tag{5.32}$$

for a given γ, where d_{ik} is the k-th element of $\mathbf{d}_i$.

Our sign-selection algorithm can thus be summarized as follows [230, 231]:

Algorithm 5.1 (CESS).

1. Let $p_0 \equiv 1$, and $p_k = 0.5$ for $k = 1, ..., N-1$. Let $\rho = 0.1$. Define $n_s = \lceil \rho n \rceil$, where $\lceil a \rceil$ represents the smallest integer that is greater than or equal to a.
2. Generate n samples of $\mathbf{d}_i$ with respect to $\mathbf{p}$, and calculate their PAR, i.e., $\mathbb{L}(\mathbf{d}_i)$.
3. Sort these samples in ascending order according to $\mathbb{L}(\mathbf{d}_i)$. Denote the obtained PAR value sequence as $[\mathbb{L}_0, ..., \mathbb{L}_{n-1}]$.
4. Find γ as $\gamma = \mathbb{L}_{n_s}$, and update $\mathbf{p}$ by using (5.32).
5. If $0 < p_k < 1$ for some k, go to step 2, and repeat this procedure by using the updated p_k.
6. Otherwise, output the optimal sign sequence as[1] $\mathbf{s}^{(\text{opt})} = 1 - 2\mathbf{p}$.

The optimization will converge to $p_k = 0$ or 1 for all k [229]. Alternatively, we may also stop the optimization after K iterations, and select the sample with the lowest PAR for transmission.

Remark 5.1. The convergence can be speeded up by considering that, if $p_k > 1 - 1/n$ (or $p_k < 1/n$), p_k will be 1 (or 0) in all n samples with high probability. However, both the numerator and the denominator in (5.32) are integers, and the denominator

[1] If all p_k are either 1 or 0, $\mathbf{p}$ can generate only one sample $\mathbf{d} = \mathbf{p}$.

is in the range[2] between n_s to n. Moreover, by observing the transition of p_k, we can see that most probably, p_k will eventually reach 1 when p_k is larger than a threshold in some stage. Therefore, we can define a threshold p_{th} in the range between $1/n_s$ and $1/n$ or larger (e.g. 0.2 for $n_s = 7$). In each iteration after updating $\mathbf{p}$ using (5.32), we let

$$p_k = \begin{cases} 1, & \text{when } p_k > 1 - p_{\text{th}}, \\ 0, & \text{when } p_k < p_{\text{th}}, \\ \text{No change}, & \text{otherwise.} \end{cases}$$

Remark 5.2. We further exploit the Tabu search [217–220] to reduce the execution time. At the last several iterations where most p_k are 1 or 0, the n samples contain many duplicated samples. Therefore, when a sample is generated, it may be with previously generated samples. If it has been previously generated, its PAR needs not be computed again. Because a list containing the samples of $\mathbf{d}$ and their PAR is also required in (5.32), we utilize this list for comparisons. Our simulation results show that about 15% of IFFTs can be saved.

5.4.3 Modified CE Sign-Selection Algorithm with Threshold

Now, we propose a Modified CE Sign-Selection Algorithm with Threshold (MCESST). This algorithm stops the optimization when the PAR of an OFDM signal is reduced to within the amplifier's linear range. Based on the Full Adaptive CE method (FACE) [227, 228], MCESST uses "elite" samples to adaptively adjust ρ. That is, instead of using ρ, we define the number of "elite" samples, n_e. After generating n samples $\mathbf{d}_1$, ..., $\mathbf{d}_n$ and sorting the corresponding $L(\mathbf{d}_i)$ in ascending order (denoting as L_1, ..., L_n), we can find a threshold γ such that $\gamma = L_{n_e}$. Then

$$\rho = P(\mathbb{L}(\mathbf{d}) \leq \gamma) \approx \hat{\rho}, \tag{5.33}$$

where $\hat{\rho} = n_e/n$. Based on the Central Limit theorem, $\hat{\rho}$ is a Gaussian random variable with mean ρ. Its variance decreases to 0 when n goes to ∞.

We can also rewrite (5.33) as

$$P(\mathbb{L}(\mathbf{d}) \leq \hat{\gamma}) \approx \frac{n_e}{n}, \tag{5.34}$$

[2]The denominator is greater than n_s when duplicated low PAR samples are generated.

where $\hat{\gamma}$ is an estimation of γ. That is, $\hat{\gamma}$ is a random variable with mean γ, and its variance decreases to 0 when n goes to ∞.

ρ determines the tradeoff between PAR reduction and execution time. A large ρ implies a large PAR reduction but a slow convergence. A small ρ leads to a fast convergence but a small PAR reduction. Therefore, we may start with a relatively large ρ (i.e., a small $n = n_{\min}$) to obtain a large PAR reduction. The slow convergence rate will not significantly increase the execution time because n is small. After each iteration, we increase n by n_{inc} until n equals a relatively large number $n_{\max}$. Because ρ decreases in each iteration with the increase of n, a fast convergence is obtained. In each iteration, if the PAR is lower than a predefined IBO threshold A, the algorithm is stopped. Our algorithm is summarized as follows:

Algorithm 5.2 (Modified CESS algorithm with threshold).

1. Initialize A, n_e, $n_{\min}$, $n_{\max}$, n_{inc}. Let $n = n_{\min}$.
2. Let $p_0 \equiv 1$, and initialize $p_k = 0.5$, $k = 0, ..., N-1$.
3. Generate n samples $\mathbf{d}_i$ with respect to $\mathbf{p}$, and calculate their PAR (i.e., $\mathbb{L}(\mathbf{d}_i)$). Whenever a $\mathbf{d}_k$ leads to PAR $\leq A$, then output $\mathbf{s}^{(\text{opt})} = 1 - 2\mathbf{d}_k$, and the algorithm ends.
4. Otherwise, sort these samples in ascending order according to $\mathbb{L}(\mathbf{d}_i)$. Denote the obtained PAR value sequence as $[\mathbb{L}_0, ..., \mathbb{L}_{N-1}]$.
5. Find γ as $\gamma = \mathbb{L}_{n_e}$, and update $\mathbf{p}$ by using (5.32).
6. If $\mathbf{p} \in \{0, 1\}^N$, then output $\mathbf{s}^{(\text{opt})} = 1 - 2\mathbf{p}$, and the algorithm ends.
7. Otherwise, if $n < n_{\max}$, let $n = n + n_{\text{inc}}$, and then, go to step 3.

5.4.4 Convergence and Complexity Discussion

5.4.4.1 Convergence of Our Algorithms

The CE convergence is proved in [229]. A basic assumption of the convergence is that the neighborhood of the optimum solution is also nearly optimum. Our optimization problem satisfies this assumption. The neighborhood in our case is defined in terms

of the Hamming distance. The PAR difference between the optimal sign sequence $\mathbf{s}^*$ and another sign sequence $\mathbf{s}_1$ within Hamming distance D from $\mathbf{s}^*$ can be bounded as

$$\left|\max_n |\hat{x}(n;\mathbf{s}^*)|^2 - \max_n |\hat{x}(n;\mathbf{s}_1)|^2\right| \leq \frac{8}{N}\left(\sum_{i=1}^{D}|X_{k_i}|\right)^2,$$

where k_i denote the position that $\mathbf{s}_1$ differs from $\mathbf{s}^*$. Consequently, when N is large, a small D always leads to near-optimal solutions.

Let the set $\mathcal{S} = \{\mathbf{s}_i : \mathbf{s}_i = 1 - 2\mathbf{d}_i, L(\mathbf{d}_i) \leq \gamma\}$, which contains n_s sign sequence samples that lead to small PARs. Intuitively, our algorithms count the number of 1's that appear in $\mathcal{S}$ to update $\mathbf{p}$. Some p_k in the updated $\mathbf{p}$ may be close to 1 (or 0). This fact implies that the signs of the corresponding subcarriers will have a large chance to be 1 (or 0) in the next iteration. Thus, most of the samples generated in the next iteration have a small Hamming distance to $\mathcal{S}$ (in the neighborhood of $\mathcal{S}$). A local optimum that has a PAR no larger than those of $\mathcal{S}$ is thus ensured. The other samples at large Hamming distances from $\mathcal{S}$ may also have small PARs, which ensures that escaping from a local optimum and converging towards the global optimal solution is possible.

5.4.4.2 Complexity of Our Algorithms for Finding a Near Optimal Solution

We first analyze the complexity of our algorithms[3] by calculating the number of samples required to find a near optimal solution.

In the $(i-1)$-th iteration, our algorithms estimate the probability $\rho = \Pr[L(\mathbf{d}) \leq \gamma]$ and optimize $\mathbf{p}$ such that, most probably, $L(\mathbf{d}) \leq \gamma$ in the i-th iteration. The accuracy of estimating ρ is determined by n_e. Therefore, we consider the complexity of finding a near optimal solution when n_e is fixed. We assume that n_e is large and ρ is small such that the samples generated in each iteration describe $\mathbf{p}$ with a negligible error.

We also assume that the optimization leads to a negligible error. Then, in the i-th iteration, no sample with $L(\mathbf{d}) > \gamma$ will be generated. In other words, the size of the search space at the i-th iteration, denoted as S_i and $S_1 = 2^N$, is reduced to

[3]To find a near optimal solution, no threshold should be used in Algorithm 5.2.

$S_i \approx \rho S_{i-1}$. If a near optimal solution is found in the K-th iteration, we have

$$S_{\min} = S_K = \rho^{K-1} S_1 = \rho^{K-1} 2^N,$$

where $S_{\min} \geq 1$ is the number of the near optimal solutions. Then,

$$K = \frac{\log_2 S_{\min} - N}{\log_2 \rho} + 1.$$

The total number of samples generated in the optimization is nK. Our simulation results show that n must be proportional to N in order to obtain the same PAR distribution for different N. Therefore, the total samples required to find a near optimal solution is of the order $\mathcal{O}(N^2)$. An exhaustive search may search 2^{N-1} samples to find the optimal solution. SLM and derandomization cannot find a near optimal solution.

5.4.4.3 Complexity Comparison for Finding Suboptimal Solutions

We may fix the number of samples or set up a threshold (as in Algorithm 5.2) to find suboptimal solutions with low complexity. Each sample is calculated by using FFT. Thus, the complexity of our algorithms for finding suboptimal solutions is $\mathcal{O}(N \log N)$ multiplications (the same as SLM). However, our simulations show that SLM requires more samples to obtain the same PAR as CESS obtains.

Derandomization iteratively calculates the signs as

$$s_k = -\text{sign}\left(\sum_{n=1}^{2JN} \sinh\left(\mu \sum_{r=1}^{k-1} s_r a_{nr}\right) \sinh(\mu a_{nk}) \prod_{r=k+1}^{N} \cosh(\mu a_{nr})\right), \tag{5.35}$$

where $\text{sign}(x)$ is the sign of x, μ is a constant, and $\{a_{nk}\}$ is the set of the real and imaginary parts of $X_k e^{j2\pi \frac{nk}{JN}}$, which are calculated and stored before calculating s_k. Computing $\{a_{nk}\}$ requires $\mathcal{O}(N^2)$ multiplications. The memory requirement for storing $\{a_{nk}\}$ is $2JN^2$ floating-point real numbers. Computing $\cosh(\mu a_{nr})$ requires $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ hyperbolic functions. Thus, calculating $\prod_{r=k+1}^{N} \cosh(\mu a_{nr})$ requires $\mathcal{O}(N^3)$ multiplications and $\mathcal{O}(N^3)$ hyperbolic functions. The rest of (5.35) requires $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ hyperbolic functions. Therefore, derandomization requires $\mathcal{O}(N^3)$ multiplications and $\mathcal{O}(N^3)$ hyperbolic functions. The execution time of the hyperbolic functions is much higher than that of multiplications.

By using more memory, we may calculate and store the results of $\prod_{r=k+1}^{N} \cosh(\mu a_{nr})$ for all k before the optimization. Then, the total memory requirement is $4JN^2$ floating-point real numbers, but the execution time is reduced to $\mathcal{O}(N^2)$ multiplications and $\mathcal{O}(N^2)$ hyperbolic functions.

A greedy algorithm is proposed in [23] based on derandomization, with which the signs are iteratively calculated as[4]

$$s_k = \arg \min_{s_k \in \{+1,-1\}} \left\| \sum_{i=1}^{k} \mathbf{a}_i s_i \right\|_p^p,$$

where $\mathbf{a}_i = [a_{ni}]$, $n = 1...2JN$ and $k = 2...N$, are vectors each having $2JN$ items. [23] proves that $p^{(\text{opt.})} = \log 2JN$. Calculating $\mathbf{a}_i$ requires $\mathcal{O}(N^2)$ multiplications. Calculating s_k, $k = 2...N$, requires $4JN(N-1)$ power-p operations[5]. Therefore, the complexity of this algorithm is $\mathcal{O}(N^2)$. Its memory requirement is $2JN^2$ floating-point real numbers (to store all $\mathbf{a}_i$).

The Algorithm 3 in [23] iteratively calculates the signs as

$$s_k = \arg \min_{s_k} \cosh \underbrace{\left(\sum_{i=1}^{k} \mathbf{a}_i s_i \right)}_{\mathbf{x}},$$

where $k = 2...N$. $\mathbf{x} = [x_n]$ is a vector of length $(2JN)$, and $\cosh \mathbf{X}$ is defined as $\sum_{n=1}^{2JN} \cosh x_n$. Therefore, it requires $\mathcal{O}(N^2)$ hyperbolic functions. Its memory requirement is also $2JN^2$ floating-point real numbers.

Derandomization and the greedy algorithm also require a large number of additions, which may be a large burden in some implementations, e.g., in using the multiplier-accumulator [116], where the execution time of the multiplications is comparable to that of the additions.

The execution time of our algorithm is smaller than that of SLM and derandomization. In the next subsection, we compare the PAR reduction of these algorithms by using simulations.

[4]When $p = \infty$, this algorithm becomes a special case of [212].

[5]The optimal p is usually large. For example, $p^{(\text{opt.})} \approx 6$ or 7 when $N = 128$ and $J = 4$.

5.4.5 Numerical Results

5.4.5.1 PAR Reduction by Using CESS

We first show the amount of PAR reduction obtained by using CESS for a 128-subcarrier, 16QAM OFDM system if large execution time is allowed. The choice of n and ρ is critical to the system performance. A large n leads to a small PAR, while the execution time increases accordingly. Similarly, a large ρ implies that many "good" samples leading to $\mathbb{L}(\mathbf{d}_i) \leq \gamma$ are taken into account in each iteration. Because CESS counts the number of 1's in these "good" samples to approximate the PAR probability distribution, more "good" samples imply more accurate estimation of a small PAR but a lower convergence rate. [227] suggests that $\rho = 0.01$ to 0.1 and $n = 3N$ to $5N$. However, such a large n may not be practical for PAR reduction. In order to find a good tradeoff between PAR reduction and execution time, we investigate the relationship between PAR reduction and the averaged number of IFFTs for different n.

Fig. 5.6 shows the averaged PAR versus the averaged number of IFFTs required by CESS for different N and n. $\rho = 0.1$ is used in this simulation. In this figure, the marks on each curve, from left to right, represent the cases of $n = 8$, 16, 32, 64, 128, 256, 512, 1024, and 2048, respectively. With $\rho = 0.1$ and $n = 8$, CESS is the same as SLM because only one sample will lead to $\mathbb{L}(\mathbf{d}_i) \leq \gamma$ in each iteration. Fig. 5.6 suggests the following:

1. All curves tend to converge to the same PAR level with large n. This finding supports the conclusion in [22] that, with optimal sign selection, the PAR can be bounded to a constant which is independent of N.

2. The averaged number of IFFTs is increased three times when n is doubled.

3. To reach the same averaged PAR for different N, n has to be $n = N/K$, where K is a constant independent of N.

4. When $n \geq 64$, further increasing n will not reduce PAR much compared to the much larger increase of execution time.

Now, let us investigate how the choice of ρ affects the averaged PAR and execution time. Fig. 5.7 shows the averaged PAR versus the averaged number of IFFTs required

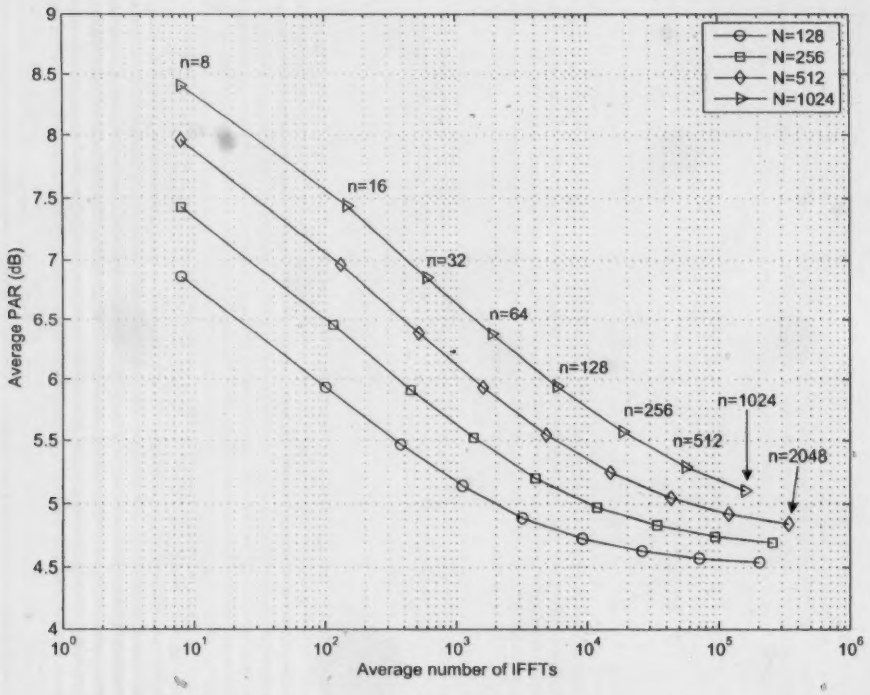


Figure 5.6: Relationship between averaged PAR and the averaged number of IFFTs required by CESS for different n, where $N = 128$ and 16QAM symbols are used.

by CESS for fixed $n = 128$ but different N and ρ. In this figure, the marks on each curve, from left to right, represent the cases of $\rho = 0.01$, 0.02, 0.04, 0.06 , 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, and 0.20, respectively. A larger ρ leads to a larger PAR reduction. $\rho = 0.1$ may be a good choice for finding a nearly optimal solution. When $\rho > 0.1$, the performance improvement in terms of the averaged PAR is small.

5.4.5.2 Performance Comparison of CESS, SLM and the Derandomization Method

Fig. 5.8 compares the PAR reduction of CESS, SLM and derandomization [22], where $N = 128$, and 16QAM symbols are used. The CESS algorithm is used with $\rho = 0.1$, and the threshold $p_{\text{th}} = 0.2$ is used for CESS with acceleration. At a 10^{-4} clip probability, CESS with $n = 30$ obtains 5 dB PAR reduction, which is 0.3 dB larger than that obtained from derandomization and 1 dB larger than that from SLM with 16 candidates. CESS with $n = 64$, which requires in average 973 sign sequence samples, obtains a 5.6 dB PAR reduction, which is about 1 dB larger than that obtained from

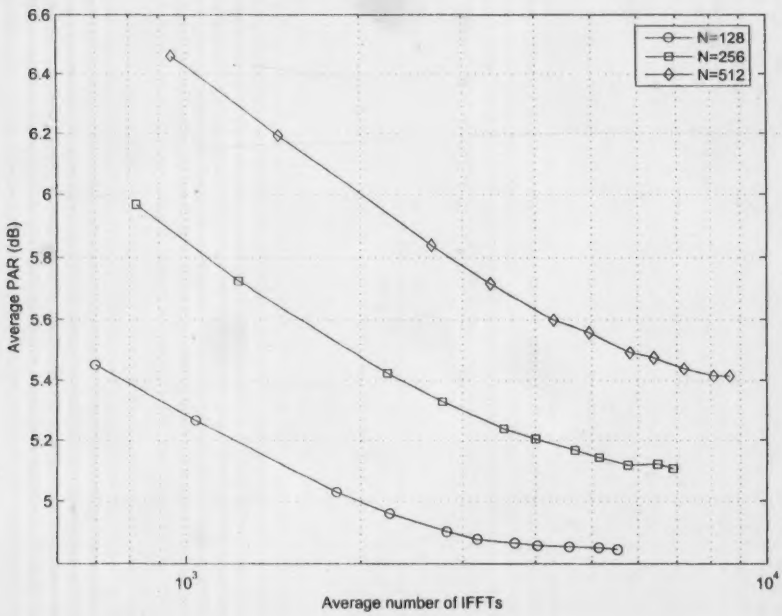


Figure 5.7: Averaged PAR v.s. the averaged number of IFFTs required by CESS for $n = 128$ and different ρ.

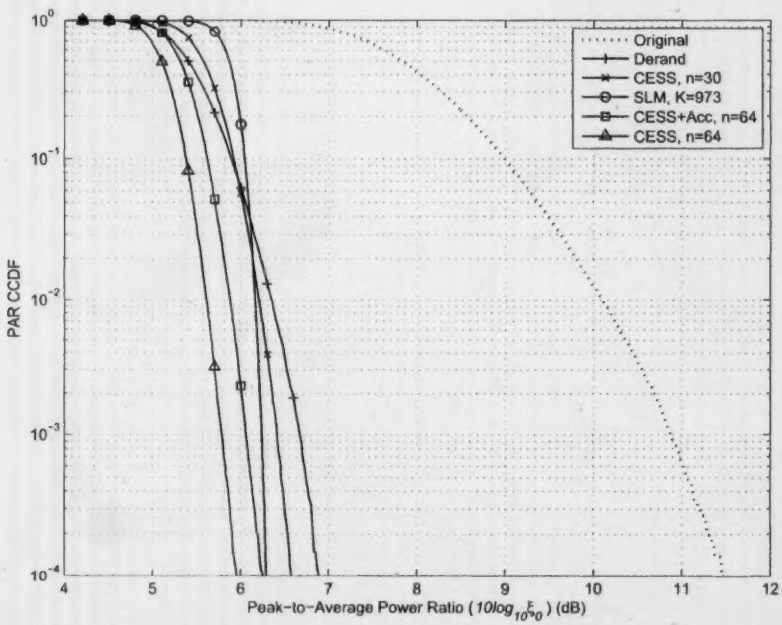


Figure 5.8: PAR reduction comparison of CESS, SLM and the derandomization method, where $N = 128$ and 16QAM symbols are used.

the derandomization method, and is 0.4 dB larger than that from SLM with 973 candidates. With convergence acceleration, the PAR reduction obtained by using CESS degrades by 0.3 dB, but its execution time is 65% of CESS without acceleration.

CESS with $n = 30$ requires an average of 235 sign sequence samples. Generating each sample requires a 512-point IFFT (equivalent to 3076 real multiplications[6] [197]). If we apply the condition (4.44) to exclude the small-magnitude samples, the execution time of calculating the PAR is negligible. Thus, the overall execution time is 722860 real multiplications.

The execution time of the derandomization method can be calculated as $(6JN^2 - 8JN)$ hyperbolic functions and $(11JN^2 - 8JN)$ real multiplications with the help of a lookup table of $4JN^2$ floating-point entries[7]. Therefore, the derandomization method requires 716800 real multiplications and 389120 hyperbolic functions with the help of a lookup table of 262144 floating-point entries to store the intermediate results used in the optimization. Thus, the execution time of derandomization is much larger than that of CESS with $n = 30$. Moreover, the memory requirement of the lookup table used by derandomization is large.

We next compare the PAR reduction of CESS and SLM for the same execution time. 16QAM symbols are used in this comparison. We consider two cases with $N = 128$ and $N = 256$, respectively. We choose $\rho = 0.1$ for CESS with $n = 8$, 16, 32, 64, 128 and 256, respectively. For each case of n, we count the averaged number of samples generated by CESS, and then use this average as the number of candidates for SLM. Fig. 5.9 illustrates this comparison. We see that, with the same execution time, CESS leads to a much smaller averaged PAR than that obtained by using SLM.

---

[6] A complex multiplication is counted as three real multiplications [192].

[7] The detailed execution time of derandomization is as follows: Calculating $a_{n,k}$: $3JN^2$ real multiplications, $\mu a_{n,k}$: $2JN^2$ real multiplications, $\cosh(\mu a_{n,k})$: $2JN(N-1)$ hyperbolic functions, $\prod \cosh(\mu a_{n,k})$: $2JN(N-2)$ real multiplications, $\sinh(\mu a_{n,k})$: $2JN(N-1)$ real multiplications, $\sinh\left(\mu \sum_{r=1}^{k-1} s_r a_{nr}\right)$: $2JN(N-2)$ hyperbolic functions, and calculating signs: $4JN(N-1)$ real multiplications.

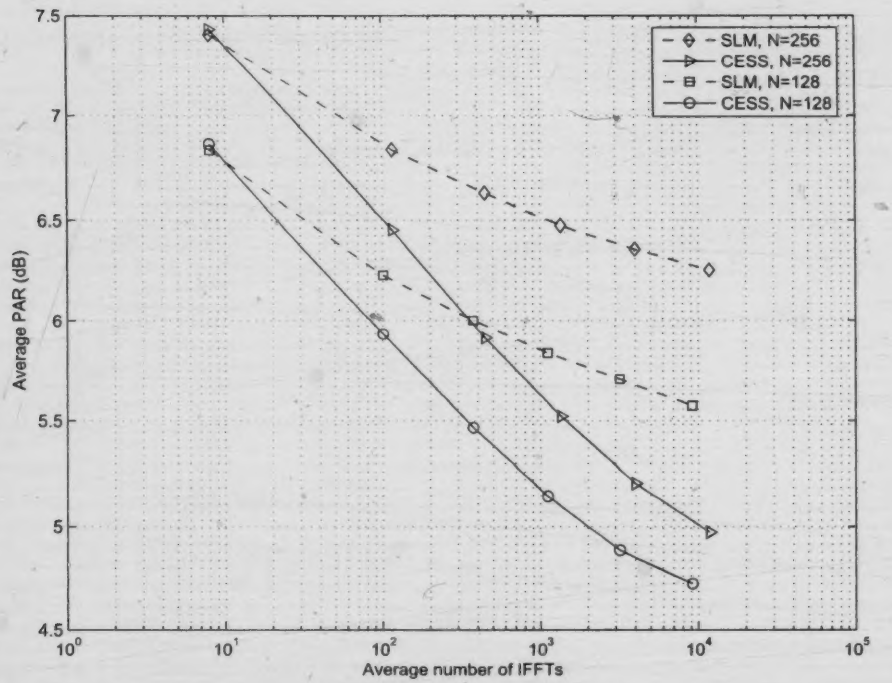


Figure 5.9: PAR reduction comparison of CESS and SLM for the same execution time, where 16QAM symbols are used.

5.4.5.3 Performance of MCESST

Now, we compare MCESST and SLM for a 128-subcarrier OFDM system. 10^5 16QAM OFDM blocks are simulated. Because the PAR of a typical single-carrier signal using square-root raised-cosine pulse shaping with a roll-off factor of 0.35 is about 6–7 dB [236], we compare MCESST and SLM with three thresholds $A = 6$ dB, 6.5 dB and 7 dB (relative to the average power). When applying a threshold to SLM, we select a large candidate set containing 10^4 sign sequence samples, but stop the SLM optimization whenever a candidate that leads to a PAR below the threshold is found. The parameters for MCESST are selected as in Table 5.2 so that the PARs of virtually all input OFDM blocks are reduced to below A. We also include in this table, out of the 10^5 simulated OFDM blocks, the numbers of "bad" OFDM blocks (denoted as "Bad Blocks") whose PARs are larger than A after optimization, as well as the worst PAR.

Fig. 5.10 shows the PAR reduction obtained by using MCESST and SLM with

Table 5.2: Simulation Parameters for MCESST

A	$n_{\min}$	$n_{\max}$	n_{inc}	n_e	Bad Blocks	Worst PAR
6.0 dB	30	80	10	4	8	6.15 dB
6.5 dB	10	350	10	3	1	6.73 dB
7.0 dB	6	18	2	3	1	7.01 dB

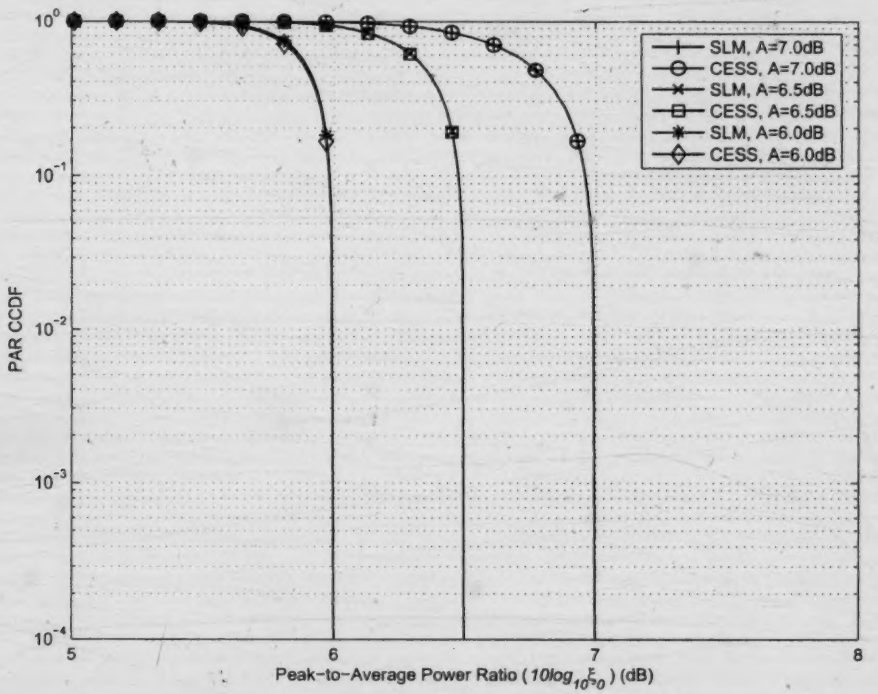


Figure 5.10: PAR reduction comparison of MCESST and SLM for different thresholds, where 16QAM symbols are used.

the above settings. Both methods obtain the same PAR reduction for each threshold used. Table 5.3 compares their execution time in terms of the averaged number of IFFTs and shows that MCESST requires less execution time than that of SLM for all threshold settings.

5.4.5.4 Performance Comparison of CESS, SLM and the Derandomization Method when Using the 91-Hex Constellation

In our previous experiments, the throughput was reduced by one bit per subcarrier. Now, we use the 91-Hex constellation [150,151] (see Section 3.3.4.2) to eliminate this

Table 5.3: Averaged Number of IFFTs required by MCESST

A	MCESST	SLM	Ratio (MCESST/SLM)
6.0 dB	196.51	556.01	35.34%
6.5 dB	33.76	40.46	83.44%
7.0 dB	7.84	7.97	98.37%

throughput loss, and we compare the PAR-reduction performance of CESS, SLM and derandomization. When using the 91-Hex constellation, on average only $\bar{N}_s = 27$ signs are allowed to change for the PAR reduction, and the averaged search space is reduced to $S = (1 + \frac{\bar{N}_s}{N})^N$.

Fig. 5.11 compares CESS with SLM and derandomization [22], where $N = 128$. The CESS algorithm is used with $\rho = 0.1$. Although only 42% of the signs are allowed to change, the PAR-reduction performance obtained by using CESS, SLM and derandomization is close to that in Fig. 5.8, where all the signs can be changed for the PAR reduction. When using the 91-Hex constellation, the PARs of these algorithms and the original OFDM signal are only a little larger than when using 16QAM.

In Fig. 5.11, CESS with $n = 30$ obtains a 5.3 dB PAR reduction at a 10^{-4} clip probability, which is 0.6 dB larger than that obtained by using the derandomization method. The averaged number of samples generated by CESS in this case is 176. With the same execution time, the PAR reduction of SLM is 0.2 dB smaller than that of CESS at a 10^{-4} clip probability, and is 0.4 dB smaller than that of CESS at a 10^{-1} clip probability. By increasing the execution time, CESS with $n = 60$ obtains a 5.8 dB PAR reduction, which is about 1.2 dB larger than that obtained by using the derandomization method.

Fig. 5.12 compares the PAR reduction obtained by using CESS and SLM for the same execution time in the 91-Hex case. As we did in the experiments in Fig. 5.9, we consider two cases with $N = 128$ and $N = 256$, respectively. We choose $\rho = 0.1$ for CESS with $n = 10$, 20, 30 and 40, respectively. Again, with the same execution time, CESS leads to a much smaller averaged PAR than that obtained by using SLM.

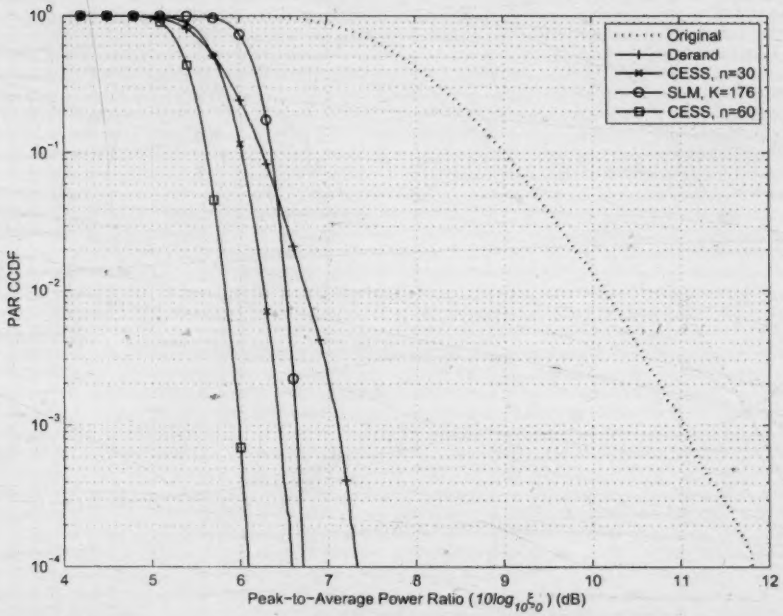


Figure 5.11: PAR reduction comparison of CESS, SLM and the derandomization method with $N = 128$ and the 91-Hex constellation.

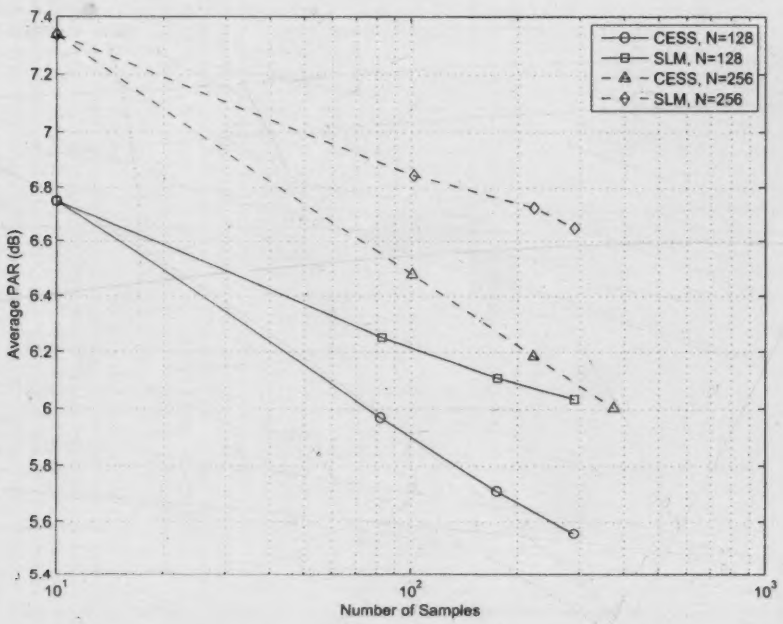


Figure 5.12: PAR reduction comparison of CESS and SLM for the same execution time with the 91-Hex constellation.

5.4.5.5 Performance of the modified CESS with threshold when Using the 91-Hex Constellation

Now, we compare MCESST and SLM for a 128-subcarrier OFDM system using the 91-Hex constellation. 10^5 OFDM blocks are simulated, and MCESST and SLM with three thresholds $A = 6\,\text{dB}$, $6.5\,\text{dB}$ and $7\,\text{dB}$ (relative to the average power) are compared. The parameters for MCESST are selected as in Table 5.4 so that the PAR of virtually all the input OFDM blocks are reduced to below A.

Table 5.4: Simulation Parameters for MCESST when 91-Hex is used

A	n_{min}	n_{max}	n_{inc}	n_{e}	Bad Blocks	Worst PAR
6 dB	30	120	10	6	33	6.15 dB
6.5 dB	10	80	10	4	5	6.54 dB
7 dB	6	18	2	4	3	7.18 dB

With these settings, MCESST and SLM obtain the same PAR reduction for each threshold used. We thus compare their execution time in terms of the averaged number of IFFTs. Table 5.5 shows the execution time comparison. MCESST requires less execution time than that of SLM for all threshold settings.

Table 5.5: Averaged Number of Samples Required by MCESST and SLM when 91-Hex is used

A	MCESST	SLM	Ratio (MCESST/SLM)
6 dB	186.94	563.93	33.15%
6.5 dB	32.15	40.66	79.07%
7 dB	7.84	7.97	98.62%

5.5 Clipping Guided Sign-Selection Algorithm for PAR Reduction

Although the optimal solution of (5.9) gives a low PAR, finding a nearly optimal solution may be computationally costly. In this section, we propose two CGS algorithms that obtain better suboptimal solutions (i.e., larger PAR reduction) of the discrete optimization problem with less execution time. These algorithms use a new clipper device to obtain the clipping noise from the time-domain OFDM signal, and then flip the signs of those subcarriers with high levels of clipping noise. These algorithms establish a connection between the discrete optimization problem and the clipping noise.

In developing these algorithms, a number of clipper models were tested. For example, we found that the conventional soft limiter [32] does not provide a sufficient performance. We thus introduce a new clipper model that requires less execution time and obtains a larger PAR reduction than those of the soft limiter. The difference between the new clipper and the soft limiter is that the clipping noise generated by the new clipper contains the entire samples of large peaks higher than a predefined threshold while that generated by the soft limiter contains only fragments of these large peaks.

We develop criteria for choosing the number of signs that should be flipped and for choosing which subcarrier signs should be flipped. Using these two criteria, we derive two sign-selection-based PAR-reduction algorithms that can handle both unitary and non-unitary signal constellations. Although a direct comparison between PTS and SLM and our CGS algorithm is not possible (as PTS and SLM use much less side information), CGS gains about 1–2 dB over these two methods for a 256-subcarrier system. A fair comparison by fixing the amount of side information is possible among CGS, derandomization and tone reservation. In this case, for a 256-subcarrier system, CGS outperforms these two methods by about 1 dB.

5.5.1 Problem Formulation

In order to derive our algorithms, we must rewrite (5.7) as

$$\hat{\mathbf{X}} = \mathbf{X} - \mathbf{C}, \tag{5.36}$$

where $\mathbf{C} = [\alpha_0 X_0, ..., \alpha_{N-1} X_{N-1}]$ is the peak-canceling vector, and

$$\alpha_k = \begin{cases} 0, & s_k = 1, \\ 2, & s_k = -1. \end{cases}$$

The discrete-time transmit signal in (5.8) may now be expressed as

$$\begin{aligned} \hat{x}_n &= \text{IDFT}[\hat{\mathbf{X}}] = x_n - c_n \\ &= x_n - \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \alpha_k X_k e^{j2\pi \frac{nk}{JN}}, \end{aligned} \tag{5.37}$$

in which c_n can be viewed as a peak-canceling signal. Only negative signs ($s_k = -1$) contribute to c_n. In the next section, we propose two suboptimal sign-selection algorithms, which utilize the strength of clipping noise on each subcarrier to find the signs that should be negative.

5.5.2 Clipping-Noise Guided Sign-Selection Algorithms

The OFDM system with CGS is shown in Fig. 5.13. The input bitstream is first mapped to data symbols X_k by using a signal constellation (or multiple constellations if bit-loading is employed). Each block of N data symbols (called an OFDM block) is fed into an IDFT block to generate the signal x_n. The CGS block finds the optimal time-domain OFDM signal $\hat{x}_n^{(\text{opt})}$ with the lowest PAR. With the CP appended to $\hat{x}_n^{(\text{opt})}$, the time-domain OFDM signal is modulated to the carrier frequency, amplified by the HPA, and transmitted through the antenna.

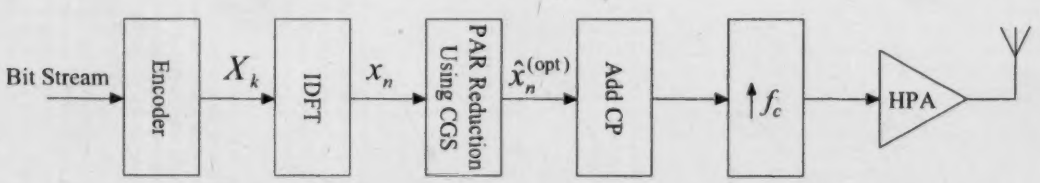


Figure 5.13: OFDM system with clipping-noise guided sign-selection.

The CGS block is expanded in Fig. 5.14, and the IDFT block is also included to facilitate the execution time calculation discussed later. In each iteration, the new clipper process samples x_n and outputs the clipping noise f_n. The DFT of the clipping noise F_k, $k = 0, ..., N-1$, are used to calculate an index set of negative signs, $\mathcal{S} = \{k : \alpha_k = 2\}$, that *may* reduce the peaks of x_n below threshold A. A

candidate OFDM block is thus generated and is used as the input of the next iteration. After L iterations, the candidate OFDM block with the smallest PAR is selected for transmission.

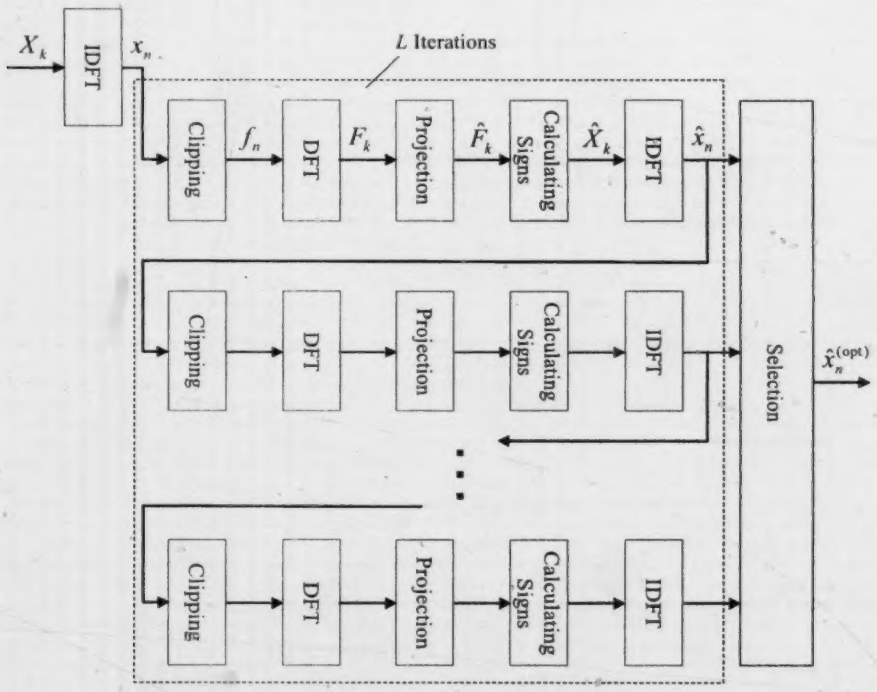


Figure 5.14: Clipping-noise guided sign-selection algorithm.

The proposed CGS algorithms are based on the new clipper model. Given a predefined threshold A, the clipping noise in this case is

$$f_n = \begin{cases} x_n, & |x_n| \geq A, \\ 0, & |x_n| < A. \end{cases} \tag{5.38}$$

This clipper is different from the conventional soft limiter that generates the clipping noise as [32]

$$f_n = \begin{cases} x_n - Ae^{j\phi_n}, & |x_n| > A, \\ 0, & |x_n| \leq A, \end{cases} \tag{5.39}$$

where ϕ_n is the phase of x_n. While both the clipper models generate the clipping noise when x_n's have large peaks, the new model (5.38) contains the *entire* samples of large peaks higher than A, whereas the soft limiter (5.39) contains only fragments of the samples which exceed A. Although both models can work with our proposed

algorithms, the new model results in less execution time. Moreover, the simulations show that the new model leads to a larger PAR reduction for non-unitary constellation input than that obtained by using the soft limiter.

In this section, we first develop two criteria for selecting signs by using the clipping noise. We propose the clipping-noise-guided sign-selection algorithm for unitary constellations (e.g., PSK), where $|X_k| \equiv 1$, and extend it to non-unitary constellations (e.g., QAM).

5.5.2.1 Sign-Selection Criteria

Recall that only the subcarriers with the negative signs contribute to the peak-canceling signal. Finding the index set of negative signs $\mathcal{S}$ requires two criteria involving the determination of the number of negative signs (which is the size of $\mathcal{S}$) and the selection of the elements of $\mathcal{S}$. Using the relationship between the peak-canceled samples and the original samples (5.37), we have

$$|\hat{x}_n| \geq |x_n| - \frac{2}{\sqrt{N}} \sum_{k \in \mathcal{S}} |X_k| .$$

Thus, a necessary condition for limiting $|\hat{x}_n|$ to smaller than A is that $\mathcal{S}$ must satisfy

$$\frac{2}{\sqrt{N}} \sum_{k \in \mathcal{S}} |X_k| \geq |x_n|_{\max} - A. \tag{5.40}$$

On the other hand, because

$$|\hat{x}_n| \leq |x_n| + \frac{2}{\sqrt{N}} \sum_{k \in \mathcal{S}} |X_k| ,$$

a larger size of $\mathcal{S}$ increases the chance of obtaining a "bad" candidate OFDM block with a large PAR. Therefore, our algorithm determines the size of $\mathcal{S}$ as follows:

Criterion 5.1 (Size of $\mathcal{S}$). *The $\mathcal{S}$ that* may *limit $|\hat{x}_n|$ to no larger than A must have the minimum size and satisfy (5.40).*

Selecting the elements of $\mathcal{S}$ depends on the clipping noise spectrum

$$F_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{JN-1} f_n e^{-j2\pi \frac{nk}{JN}}, \quad k = 0, ..., N-1, \tag{5.41}$$

where the clipping noise f_n is given in (5.38). Projecting F_k to X_k, we have

$$\hat{F}_k = D_k X_k, \quad k = 0, ..., N-1,$$

where

$$D_k = \Re\left[\frac{F_k X_k^*}{|X_k|^2}\right], \tag{5.42}$$

and $\Re[x]$ represents the real part of x, and $(\cdot)^*$ represents the complex conjugate.

A peak-canceling signal may be obtained by taking the IDFT of $\hat{F}_k$ as

$$\hat{f}_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{F}_k e^{j2\pi \frac{nk}{JN}} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} D_k X_k e^{j2\pi \frac{nk}{JN}}. \tag{5.43}$$

If a large peak of f_n, or equivalently x_n, occurs at n_0, then, $\hat{f}_n$ also has a large peak at n_0 with the same direction of f_{n_0}. Moreover, due to peak regrowth [19], usually $\left|\hat{f}_{n_0}\right| < |f_{n_0}|$. Therefore, $\hat{f}_n$ may be scaled by the optimum factor $\beta^{(\text{opt})} > 0$ to further reduce the peaks of x_n. The modified discrete samples of the OFDM signal is

$$\bar{x}_n = x_n - \beta^{(\text{opt})} \hat{f}_n = x_n - \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \beta^{(\text{opt})} D_k X_k e^{j2\pi \frac{nk}{JN}}. \tag{5.44}$$

The comparison of (5.37) and (5.44) suggests that α_k can be obtained by rounding some $\beta^{(\text{opt})} D_k$ to 2 and rounding others to 0. The resulting modified discrete samples $\hat{x}_n$ may have larger peaks than A. However, by minimizing the rounding error, the peaks of $\hat{x}_n$ may still be lower than $|x_n|_{\max}$. The mean squared rounding error is upper-bounded as

$$\begin{aligned}
\varepsilon &= \mathrm{E}\left\{|\hat{x}_n - \bar{x}_n|^2\right\} \\
&= \frac{1}{JN} \sum_n \left| \frac{1}{\sqrt{N}} \sum_{k \in \mathcal{S}} \left(2X_k - \beta^{(\text{opt})} D_k X_k\right) e^{j2\pi \frac{nk}{JN}} \right. \\
&\quad \left. + \frac{1}{\sqrt{N}} \sum_{k \notin \mathcal{S}} \beta^{(\text{opt})} D_k X_k e^{j2\pi \frac{nk}{JN}} \right|^2 \\
&\leq \frac{1}{N} \left(\sum_{k \in \mathcal{S}} \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| + \sum_{k \notin \mathcal{S}} \left|\beta^{(\text{opt})} D_k X_k\right| \right)^2.
\end{aligned} \tag{5.45}$$

The upper-bound of ε is determined by the term inside $(\cdot)^2$. Then, we have

Criterion 5.2 (Elements of $\mathcal{S}$). *Define the maximum rounding error as*

$$\varepsilon_{max} = \sum_{k \in \mathcal{S}} \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| + \sum_{k \notin \mathcal{S}} \left|\beta^{(\text{opt})} D_k X_k\right|. \tag{5.46}$$

Then, $\mathcal{S}$ must be generated by rounding $\beta^{(\text{opt})} D_k$ to 2 or 0 such that ε_{max} is minimized.

5.5.2.2 The Clipping-Noise Guided Sign-Selection Algorithm for Unitary Modulations

By substituting $|X_k| \equiv 1$ into (5.40), the size of $\mathcal{S}$ can be calculated as

$$I = \left\lceil \frac{\sqrt{N}(|x_n|_{\max} - A)}{2} \right\rceil. \tag{5.47}$$

where $\lceil x \rceil$ represents the smallest integer greater than x. In unitary modulations, minimizing $\varepsilon_{\max}$ involves D_k only.

Theorem 5.1. *For unitary modulations, ε_{max} is minimized if $D_k \geq D_m$ for all $k \in \mathcal{S}$ and $m \notin \mathcal{S}$.*

Proof. By substituting $|X_k| \equiv 1$ into (5.46),

$$\varepsilon_{\max} = \sum_{k \in \mathcal{S}} \left|2 - \beta^{(\text{opt})} D_k\right| + \sum_{k \notin \mathcal{S}} \left|\beta^{(\text{opt})} D_k\right|.$$

Let $\mathcal{S} = \{k_1, k_2, ..., k_I\}$ where $D_k \geq D_m$ for any $k \in \mathcal{S}$ and $m \notin \mathcal{S}$. Without loss of generality, let $\mathcal{S}' = \{k_1, ..., k_{I-1}, m\}$, $m \notin \mathcal{S}$. Then,

$$\begin{aligned} \varepsilon_{\max} - \varepsilon'_{\max} = &\left|2 - \beta^{(\text{opt})} D_{k_I}\right| + \left|\beta^{(\text{opt})} D_m\right| \\ &- \left|2 - \beta^{(\text{opt})} D_m\right| - \left|\beta^{(\text{opt})} D_{k_I}\right|, \end{aligned}$$

where $\varepsilon_{\max}$ and $\varepsilon'_{\max}$ are the maximum rounding errors associated with $\mathcal{S}$ and $\mathcal{S}'$, respectively. By discussing the signs of $\beta^{(\text{opt})} D_{k_I}$, $\beta^{(\text{opt})} D_m$, $(2 - \beta^{(\text{opt})} D_{k_I})$ and $(2 - \beta^{(\text{opt})} D_m)$, we have $\varepsilon_{\max} - \varepsilon'_{\max} \leq 0$. Because $m \notin \mathcal{S}$ is arbitrary, Theorem 5.1 is proved. □

Note that $\beta^{(\text{opt})}$ is irrelevant to the decision about $\mathcal{S}$ in the minimization of $\varepsilon_{\max}$ in unitary modulations. In other words, we may substitute $\beta^{(\text{opt})}$ with an arbitrary positive number without making a wrong decision about $\mathcal{S}$. However, in non-unitary modulations, we must find $\beta^{(\text{opt})}$ before we properly choose $\mathcal{S}$. We will discuss this process in the next subsection.

Also, $\varepsilon_{\max} - \varepsilon'_{\max} = 0$ when $\beta^{(\text{opt})} D_{k_I} \geq \beta^{(\text{opt})} D_m \geq 2$ or $D_m \leq D_{k_I} \leq 0$. Thus, we are free to include[8] any k to $\mathcal{S}$ if $\beta^{(\text{opt})} D_k \geq 2$ and to exclude any m from $\mathcal{S}$ if $D_m \leq 0$.

[8] We do not exploit this freedom in Algorithm 5.3 because we do not calculate $\beta^{(\text{opt})}$.

We must round the largest I samples of D_k to 2 and round others to 0. Our algorithm is then summarized as follows [237]:

Algorithm 5.3 (Clipping-Noise Guided Sign-Selection for Unitary Constellations).

Initialization:

1. Choose a magnitude threshold A, and the number of iterations L.

Runtime:

1. For each $\mathbf{X}$, calculate x_n by using (5.2). Note that $J \geq 4$ is required. Let $\eta = |x_n|_{\max}$.
2. If $\eta > A$, let the iteration number $l = 1$ and go to Step 3; otherwise, transmit x_n and terminate.
3. Set the index set of negative signs empty $\mathcal{S} = \emptyset$; calculate f_n by using (5.38).
4. Calculate D_k by using (5.42). If all $D_k \leq 0$, go to Step 8.
5. Otherwise, calculate I by using (5.47); generate $\mathcal{S}$ by rounding the largest I samples of D_k to 2 and rounding other D_k to 0.
6. Calculate $\hat{\mathbf{X}}$ and $\hat{x}_n$ by using (5.36) and (5.37), respectively. If $|\hat{x}_n|_{\max} < \eta$, let $\eta = |\hat{x}_n|_{\max}$, and store $\hat{x}_n$ as $\hat{x}_n^{(\text{opt})}$.
7. If $l = L$ go to Step 8; otherwise, increase l by one, let $\mathbf{X} = \hat{\mathbf{X}}$ and $x_n = \hat{x}_n$, and go to Step 3.
8. Transmit $\hat{x}_n^{(\text{opt})}$.

The execution time of this algorithm may be upper bounded as two FFTs per iteration. The detailed complexity analysis is given in the next subsection.

5.5.2.3 The Clipping-Noise Guided Sign-Selection Algorithm for Non-Unitary Modulations

Without loss of generality, we set $\mathrm{E}\left\{|X_k|^2\right\} = 1$. Let $\bar{d} = \mathrm{E}\left\{|X_k|\right\}$. The size of $\mathcal{S}$ can be calculated from (5.40) as

$$I = \left\lceil \frac{\sqrt{N}(|x_n|_{\max} - A)}{2\bar{d}} \right\rceil. \tag{5.48}$$

For square M-ary QAM constellations, we have

$$\bar{d} = \frac{4}{M}\sqrt{\frac{6}{M-1}} \sum_{m=1}^{\sqrt{M}/2} \sum_{n=1}^{\sqrt{M}/2} \sqrt{(m-0.5)^2 + (n-0.5)^2}. \tag{5.49}$$

In non-unitary modulations, minimizing $\varepsilon_{\max}$ requires a knowledge of $\beta^{(\text{opt})}$. By using the adaptive-scaling algorithm [103], $\beta^{(\text{opt})}$ can be found as

$$\beta^{(\text{opt})} = \frac{\Re\left[\sum_{n\in\mathcal{S}_p} f_n \hat{f}_n^*\right]}{\sum_{n\in\mathcal{S}_p} |\hat{f}_n|^2}, \tag{5.50}$$

where $\mathcal{S}_p = \{n : n \in \mathcal{S}_1, |x_n| > |x_{n-1}|, \text{ and } |x_n| \geq |x_{n+1}|\}$ is the index set of the peaks of f_n.

Because $|X_k|$ is not a constant, the minimization of $\varepsilon_{\max}$ for non-unitary modulations depends on both $\beta^{(\text{opt})} D_k$ and $|X_k|$. To illustrate the relationship between $\beta^{(\text{opt})} D_k$, $|X_k|$ and $\varepsilon_{\max}$, we consider the following example:

Example 5.1. Suppose we have chosen $I-1$ elements of $\mathcal{S}$, and the last element will be selected from $\beta^{(\text{opt})} D_0$ and $\beta^{(\text{opt})} D_1$,where $\beta^{(\text{opt})} D_0 = 1.1$ and $\beta^{(\text{opt})} D_1 = 1.04$.

Case I: X_k are 4QAM symbols.

Because $D_0 > D_1$, $\varepsilon_{\max}$ is minimized when $\alpha_0 = 2$ and $\alpha_1 = 0$.

Case II: X_k are 16QAM symbols, and $X_0 = 0.5 + 0.5j$, $X_1 = 1.5 + 1.5j$.

It is easy to verify that $\varepsilon_{\max} = 2.84$ if $\alpha_0 = 2$ and $\alpha_1 = 0$. However, $\varepsilon_{\max} = 2.81$ when $\alpha_0 = 0$ and $\alpha_1 = 2$. Thus, the optimum sign-selection is $\alpha_0 = 0$ and $\alpha_1 = 2$ although $D_0 > D_1$. □

In general, we have the following theorem for minimizing $\varepsilon_{\max}$ in non-unitary cases.

Theorem 5.2. *In each iteration, let $\mathcal{S}$ have the size I that is calculated in (5.48). Define*

$$T_k = \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| - \left|\beta^{(\text{opt})} D_k X_k\right|. \tag{5.51}$$

Then, for non-unitary modulations, ε_{max} is minimized if

$$T_k \leq T_m \text{ for all } k \in \mathcal{S} \text{ and } m \notin \mathcal{S}. \tag{5.52}$$

Proof. Suppose $\mathcal{S}$ satisfies (5.52). Let $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ and $\mathcal{S}_1 \cap \mathcal{S}_2 = \emptyset$. We may form another set $\mathcal{S}'$ by replacing $\mathcal{S}_2$ with $\mathcal{S}_3$ where $\mathcal{S}_3 \cap \mathcal{S} = \emptyset$. Because $\mathcal{S}$ and $\mathcal{S}'$ have the same size, $\mathcal{S}_2$ and $\mathcal{S}_3$ also have the same size. In the following, we show that $\mathcal{S}$ leads to a smaller maximum rounding error than $\mathcal{S}'$.

Let $\mathcal{S}_4$ contain the indices that are not in $\mathcal{S}_1$, $\mathcal{S}_2$ and $\mathcal{S}_3$; i.e.,

$$\bigcup_{i=1}^{4} \mathcal{S}_i = \mathcal{N},$$

where $\mathcal{N} = [0, ..., N-1]$ is the OFDM index set, and

$$\mathcal{S}_i \cap \mathcal{S}_k = \emptyset, \quad \text{for any } i \neq k.$$

The maximum rounding error caused by $\mathcal{S}$ is then given by

$$\varepsilon_{\max} = \sum_{k \in \mathcal{S}_1 \cup \mathcal{S}_2} \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| + \sum_{k \in \mathcal{S}_3 \cup \mathcal{S}_4} \left|\beta^{(\text{opt})} D_k X_k\right|,$$

and the maximum rounding error caused by $\mathcal{S}'$ is

$$\varepsilon'_{\max} = \sum_{k \in \mathcal{S}_1 \cup \mathcal{S}_3} \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| + \sum_{k \in \mathcal{S}_2 \cup \mathcal{S}_4} \left|\beta^{(\text{opt})} D_k X_k\right|.$$

Therefore, we have

$$\begin{aligned}
\varepsilon_{\max} - \varepsilon'_{\max} &= \sum_{k \in \mathcal{S}_2} \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| + \sum_{k \in \mathcal{S}_3} \left|\beta^{(\text{opt})} D_k X_k\right| \\
&\quad - \sum_{k \in \mathcal{S}_3} \left|(2 - \beta^{(\text{opt})} D_k) X_k\right| - \sum_{k \in \mathcal{S}_2} \left|\beta^{(\text{opt})} D_k X_k\right| \\
&= \sum_{k \in \mathcal{S}_2} \left(\left|(2 - \beta^{(\text{opt})} D_k) X_k\right| - \left|\beta^{(\text{opt})} D_k) X_k\right|\right) \\
&\quad - \sum_{k \in \mathcal{S}_3} \left(\left|(2 - \beta^{(\text{opt})} D_k) X_k\right| - \left|\beta^{(\text{opt})} D_k X_k\right|\right) \\
&= \sum_{k \in \mathcal{S}_2} T_k - \sum_{k \in \mathcal{S}_3} T_k \leq 0.
\end{aligned}$$

Because $\mathcal{S}_2$ and $\mathcal{S}_3$ are arbitrarily selected, $\mathcal{S}$ minimizes the maximum rounding error.

□

Our algorithm for non-unitary modulations can be modified from Algorithm 5.3 as follows [237] (the omitted parts are the same as those in Algorithm 5.3).

Algorithm 5.4 (Clipping Guided Sign-Selection for Non-Unitary Modulations).

• • •

5. Find $\mathcal{S}$ as follows:

   5-a. Calculate I, $\beta^{(\text{opt})}$ and T_k by using (5.48), (5.50) and (5.51), respectively.

   5-b. Find the smallest I samples of T_k, and denote them as $T_{k_1}, ..., T_{k_I}$; set $\alpha_i = 2$ for $i = k_1, ..., k_I$, and set other α_i to 0.

• • •

Because calculating $\beta^{(\text{opt})}$ involves an FFT (for calculating $\hat{f}_n$), the execution time of this algorithm is upper-bounded as three FFTs per iteration.

Remark 5.3 (Adaptively calculating the size of $\mathcal{S}$). We may directly use (5.40) to find the size of $\mathcal{S}$. That is, in each iteration, we start from an empty set $\mathcal{S} = \emptyset$ and flip the signs one-by-one. The condition in (5.40) is tested every time a sign is flipped. When (5.40) is satisfied, we generate the $\hat{\mathbf{X}}$ and go to the next iteration.

The size of $\mathcal{S}$ now depends on the choice of its elements. Therefore, the condition of Theorem 5.2 is not satisfied. Although in this case, (5.52) does not ensure the minimization of $\varepsilon_{\max}$, the resulting $\varepsilon_{\max}$ is still small. Moreover, an adaptive size of $\mathcal{S}$ more precisely meets Criterion 5.1. Thus, adaptively calculating the size of $\mathcal{S}$ will not degrade the PAR-reduction performance. In fact, our simulations show that the PAR-reduction performance is improved by using an adaptive size of $\mathcal{S}$.

Remark 5.4 (Simplification). The execution time of Algorithm 5.4 can be simplified to two FFTs per iteration by using the mean of $\beta^{(\text{opt})}$, which is calculated at the initialization stage [16], and is used for all OFDM blocks.

Unless A is small, the clipping noise consists of a series of pulses [19]. In [19], we have proved that, if the clipping noise contains only one dominant pulse, the peak of $\hat{f}_n$ is proportional to that of f_n:

$$\left|\hat{f}_n\right|_{\max} = \gamma \left|f_n\right|_{\max},$$

where γ is the constant of proportionality, whose value depends on the OFDM bandwidth and the clipping level. The mean of γ is

$$\bar{\gamma} = \frac{2\sqrt{2}}{\sqrt{3\pi}} \frac{1}{A/\sigma},$$

where $2\sigma^2$ is the average power of the OFDM signals. Because $\beta^{(\text{opt})}\hat{f}_n$ must be close to f_n at the positions of most peaks of f_n, we may estimate the mean of $\beta^{(\text{opt})}$ as

$$\bar{\beta} = \frac{1}{\bar{\gamma}} = \frac{\sqrt{3\pi}A}{2\sqrt{2}\sigma}. \tag{5.53}$$

Remark 5.5 (Choice of A). As with other clipping-based PAR-reduction algorithms, the choice of A relies mainly on experiments. In general, A should be small if a large PAR reduction is required, and A should be relatively large if fast execution is desired.

Remark 5.6 (Further Discussion). Our algorithm is effective for large N and for OFDM symbols with a large PAR. This finding can be intuitively explained as follows:

1. Because SLM with a small number of candidates can effectively reduce the PAR, when N is large, a large number of sign sequences must exist that can effectively reduce the PAR. Consequently, the sign sequence obtained from rounding the clipping noise has a high probability of belonging to these sign sequences.

2. The PAR-reduction performance of our algorithms is determined mainly by the errors in estimating the size of $\mathcal{S}$ and in rounding $\beta^{(\text{opt})}D_k$ to 2 or 0. In the next section, we will see that using the average size of $\mathcal{S}$ leads to smaller PAR reduction than using the adaptive size of $\mathcal{S}$, while using $\beta^{(\text{opt})}$ or $\bar{\beta}$ gives virtually the same PAR reduction. Therefore, the accuracy of estimating the size of $\mathcal{S}$ plays a more important role in our algorithm than in the minimization of the rounding error.

   a) Effect of the error of estimating the size of $\mathcal{S}$.

      If such an estimation error occurs, we either have rounded too many D_k's to 2 or rounded too many D_k's to 0. In either case, the maximum degradation of peak reduction is

      $$\varepsilon_d = \sum_{k \in \mathcal{S}_e} \frac{2\,|X_k|}{\sqrt{N}},$$

      where $\mathcal{S}_e$ is the difference between the $\mathcal{S}$ we are using and the optimum $\mathcal{S}$. Thus, $\mathcal{S}_e$ is small when N is large.

   b) Effect of the error of rounding $\beta^{(\text{opt})}D_k$ to 2 or 0.

Ideally, a peak canceling signal is a series of pulses, which cancels the large peaks in the OFDM signal without introducing any new peaks or increasing any small peaks. Such an ideal signal does not exist because of the spectrum constraints on the peak canceling signals.

In our algorithms, the rounding error adds new peaks to the OFDM signal. In OFDM signals with small PARs, most peaks have comparable magnitudes. Thus, the new peaks introduced by the rounding error may easily fall on some existing peaks of the OFDM signal. Little PAR reduction can be obtained.

On the other hand, in OFDM signals with large PARs, because the number of large peaks is small [103], the chance that the new peaks introduced by the rounding error fall on the large peaks of the OFDM signal is also small. Thus, the large PAR can be effectively reduced.

5.5.3 Complexity Analysis

We analyze the complexity of Algorithm 5.3. The complexity of Algorithm 5.4 and its simplification can be analyzed similarly.

The complexity of our algorithm[9] is determined by mainly (5.37) and (5.41), where an IDFT and a DFT, respectively, are used. The complexity of the other calculations involved in our algorithm is $\mathcal{O}(N)$ and is independent of J. Therefore, the complexity of our algorithm is $\mathcal{O}(N \log N)$. However, because the inputs of the IDFT/DFT in (5.37) and (5.41) are sparse, their execution time is much smaller than the length-(JN) IFFT/FFT.

In Eq. (5.41), the averaged number of nonzero samples in f_n can be calculated as [103]

$$\bar{N}_f = JNe^{-A^2/(2\sigma^2)},$$

where $2\sigma^2 = 1$ is the mean power of the OFDM signal. $\bar{N}_f/N$ is usually small; e.g., $\bar{N}_f/N \approx 32\%$ for $A = 4\,\text{dB}$ and $J = 4$. Because most samples of f_n are 0, the execution time of (5.41) is much lower than that of a length-(JN) FFT.

[9]Here, the Runtime Step 1 is excluded because it is required in all OFDM systems whether or not PAR-reduction techniques are involved.

Eq. (5.37) also involves a small-size $\mathcal{S}$. The average size of $\mathcal{S}$ can be calculated as

$$\mathrm{E}\{I\} = \int_A^\infty I(r)p(r)dr,$$

where $I(r)$ is the size of $\mathcal{S}$ for $r = |x_n|_{\max}$,

$$I(r) = \left\lceil \frac{\sqrt{N}(r-A)}{2\bar{d}} \right\rceil,$$

and $\bar{d} = \mathrm{E}\{|X_k|\}$. By using the approximated PAR cumulative distribution function (CDF) [36],

$$\Pr(\max|x_n| < r) \approx (1 - e^{-r^2})^{\alpha N},$$

where α is empirically obtained as 2.8, the probability density function (pdf) of PAR can be found as

$$p(r) = 2\alpha N r e^{-r^2}(1 - e^{-r^2})^{\alpha N - 1}.$$

Then,

$$\mathrm{E}\{I\} = \int_A^\infty \left\lceil \frac{\sqrt{N}(r-A)}{2\bar{d}} \right\rceil p(r)dr.$$

When $N = 256$ and $A = 4\,\mathrm{dB}$, $\mathrm{E}\{I\} \approx 8.1$ for 4QAM and $\mathrm{E}\{I\} \approx 8.6$ for 16QAM, where the simulated results are 7.9 and 8.4, respectively. If the adaptive size of $\mathcal{S}$ is used, $\mathrm{E}\{I\}$ will be a little larger. For example, for $N = 256$, $A = 4\,\mathrm{dB}$ and 16QAM symbol input, the mean of the adaptive size of $\mathcal{S}$ obtained by simulation is 12.6. In any case, the execution time of (5.37) is much less than that of a length-(JN) IFFT.

To simplify the execution time comparison, we loosely upper bound the execution time of Algorithm 5.3 as two FFTs per iteration. With similar analysis, we can show that the execution time upper-bounds of Algorithm 5.4 and its simplification are three and two FFTs per iteration, respectively.

5.5.4 Simulation Results

In this section, we compare the CGS algorithm with SLM, PTS, derandomization, and tone reservation techniques for a 256-subcarrier OFDM system with unit-energy 64QAM. Four times oversampling ($J = 4$) is used in optimization and eight times oversampling ($J = 8$) is used in calculating the PAR after an optimal sign sequence is selected. For CGS, the average $\beta^{(\mathrm{opt})}$, $\bar{\beta}$, and the adaptively calculated size of $\mathcal{S}$

are used (see Remarks 5.3 and 5.4). The adaptive scaling algorithm [103] with eight iterations is used for tone reservation, where randomly-selected 1/6 subcarriers are reserved for PAR reduction. Thus, the amount of redundancy is the same in tone reservation and CGS. The execution time of the tone reservation is the same as that of CGS with eight iterations.

Fig. 5.15 compares the PAR reduction of these algorithms in terms of the PAR complementary cumulative distribution function (CCDF) $F(\xi_0) = \Pr[\xi > \xi_0]$. The PAR reduction of the tone reservation is about 0.1 dB larger than that of CGS with eight iterations at $F(\xi_0) = 10^{-3}$, and is the same at $F(\xi_0) = 10^{-4}$ (about 5.9 dB). However, tone reservation leads to a 0.42 dB average power increase, which translates into a BER increase, while CGS does not change the average power.

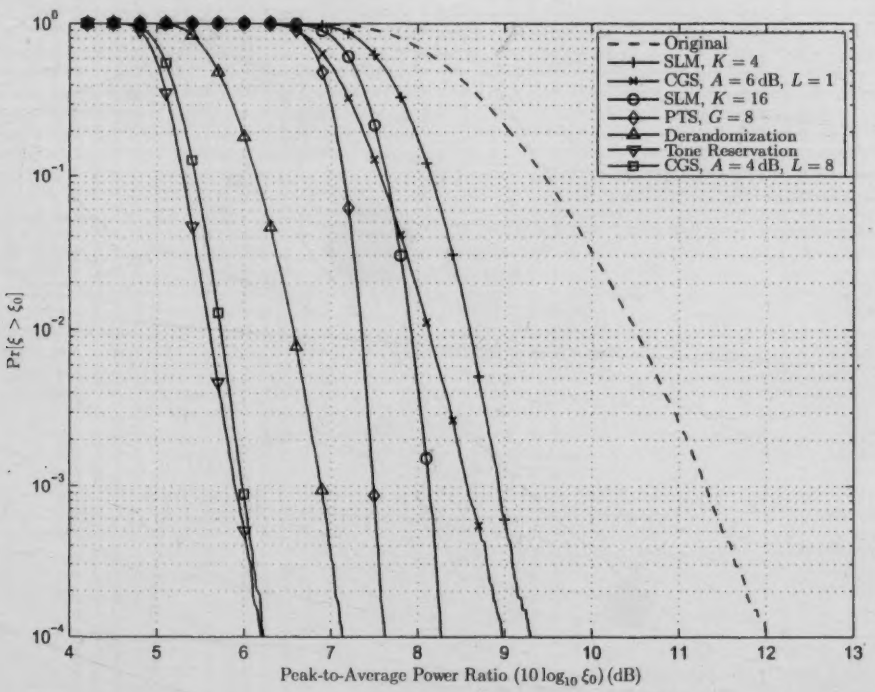

Original
SLM, $K = 4$
CGS, $A = 6$ dB, $L = 1$
SLM, $K = 16$
PTS, $G = 8$
Derandomization
Tone Reservation
CGS, $A = 4$ dB, $L = 8$
$\Pr[\xi > \xi_0]$
Peak-to-Average Power Ratio $(10 \log_{10} \xi_0)$ (dB)

Figure 5.15: PAR reduction comparison of CGS with $\bar{\beta}$ and adaptive size of $\mathcal{S}$, SLM, PTS, derandomization and tone reservation, where $N = 256$, and 64QAM symbol input is used.

At $F(\xi_0) = 10^{-4}$, the PAR reduction of CGS with eight iterations is about 2.1 dB larger than that of SLM with 16 candidates, 1.4 dB larger than that of PTS with eight randomly partitioned subgroups, and 0.9 dB larger than that of the derandomization

algorithm. In this experiment, both SLM and CGS require 16 FFTs per OFDM block. PTS must test $2^7 = 128$ combinations.

CGS with one iteration obtains a 3.2 dB PAR reduction, which is about 0.3 dB larger than that of SLM with four candidates. In this experiment, CGS and SLM require two and four FFTs per OFDM block, respectively.

Table 5.6 lists the averaged running time of these algorithms obtained on a Pentium IV 3.40G computer using Matlab R14 Service Pack 2. The time used by CGS with eight iterations is only 89% of that of SLM with 16 candidates, 9.4% of that of PTS with eight subgroups, and 1.4% of that of the derandomization algorithm.

Table 5.6: Averaged running time of CGS, SLM, PTS and the derandomization algorithm, where $N = 256$, and 64QAM symbol input is used.

	Averaged Computation Time (millisecond)
CGS, $A = 4$ dB, $L = 8$	4.9
SLM, $K = 16$	5.5
PTS, $G = 8$	52.1
Derandomization	362.0

We now compare SLM, PTS, derandomization, and CGS by observing the Power Spectrum Density (PSD) of the output of the OFDM transmitter, where the peak-reduced OFDM signal is passed through a Solid State Power Amplifier (SSPA) [1],

$$y(t) = \frac{|x(t)|}{\left(1 + \left(\frac{|x(t)|}{C}\right)^{2p}\right)^{\frac{1}{2p}}} e^{j\phi(t)},$$

where $x(t) = |x(t)|e^{j\phi(t)}$ is the input, and $y(t)$ is the output of SSPA. Usually, $p = 3$ for practical SSPA. In our simulations, we choose $C = 6$ dB. Fig. 5.16 shows the result. CGS with eight iterations leads to only -40 dB out-of-band radiation, which is 1 dB lower than that of derandomization, 8 dB lower than that of PTS with eight subgroups, 9 dB lower than that of SLM with 16 candidates, and 12 dB lower than that without using any PAR-reduction techniques, respectively.

Fig. 5.17 compares Algorithm 5.4 with different configurations including $\bar{\beta}$ + average size of $\mathcal{S}$, $\bar{\beta}$ + adaptive size of $\mathcal{S}$, $\beta^{(\text{opt})}$ + average size of $\mathcal{S}$, and $\beta^{(\text{opt})}$ + adaptive

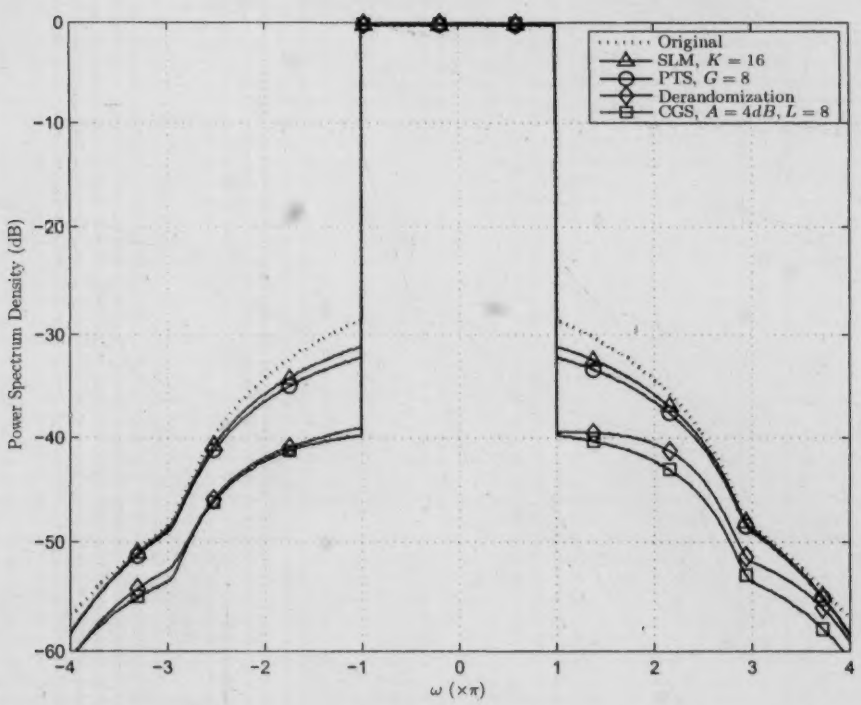


Figure 5.16: Power spectrum density of SLM, PTS, derandomization and CGS, where $N = 256$, and 64QAM symbol input is used.

size of $\mathcal{S}$. 256 subcarriers and 16QAM symbols are used. Using the adaptive size of $\mathcal{S}$ leads to larger PAR reduction than using the average size of $\mathcal{S}$. On the other hand, using $\beta^{(\text{opt})}$ leads to virtually the same PAR reduction as using $\bar{\beta}$. The execution time is three FFTs per iteration when $\beta^{(\text{opt})}$ is used while the execution time is two FFTs per iteration when $\bar{\beta}$ is used.

Fig. 5.18 compares Algorithm 5.4 with eight iterations but different clipping level A. 256 subcarriers and 16QAM symbols are used, and the CGS configuration is $\bar{\beta}$ + adaptive size of $\mathcal{S}$. $A = 4\,\text{dB}$ leads to the largest PAR reduction. Other choices of A lead to performance degradation. However, the largest performance degradation is only about 0.9 dB. Therefore, our algorithm is not overly sensitive to the choice of A.

Fig. 5.19 compares CGS with $\bar{\beta}$ + adaptive size of $\mathcal{S}$ and SLM for OFDM systems with different numbers of subcarriers, where the other parameters are same. When the number of subcarriers doubles, the PAR reduction of both algorithms decreases by about 0.4 dB.

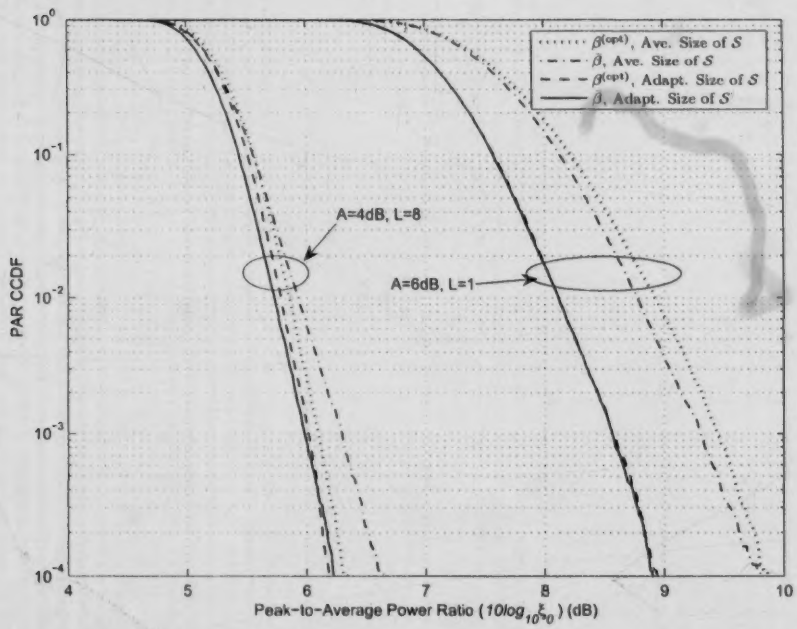


Figure 5.17: PAR reduction comparison of different configurations of CGS, where $N = 256$, and 16QAM symbol input is used.

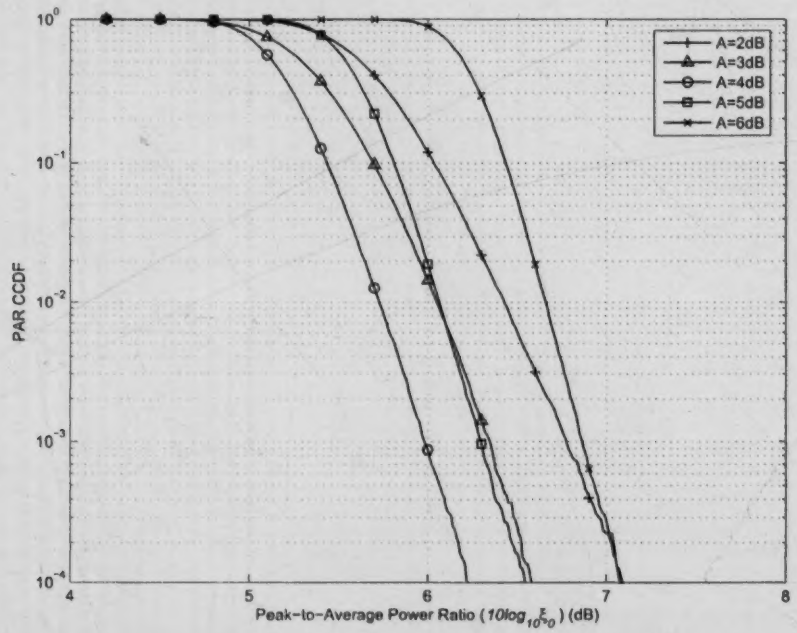


Figure 5.18: PAR reduction comparison of CGS with $L = 8$ and different A, where $N = 256$, and 16QAM symbol input is used.

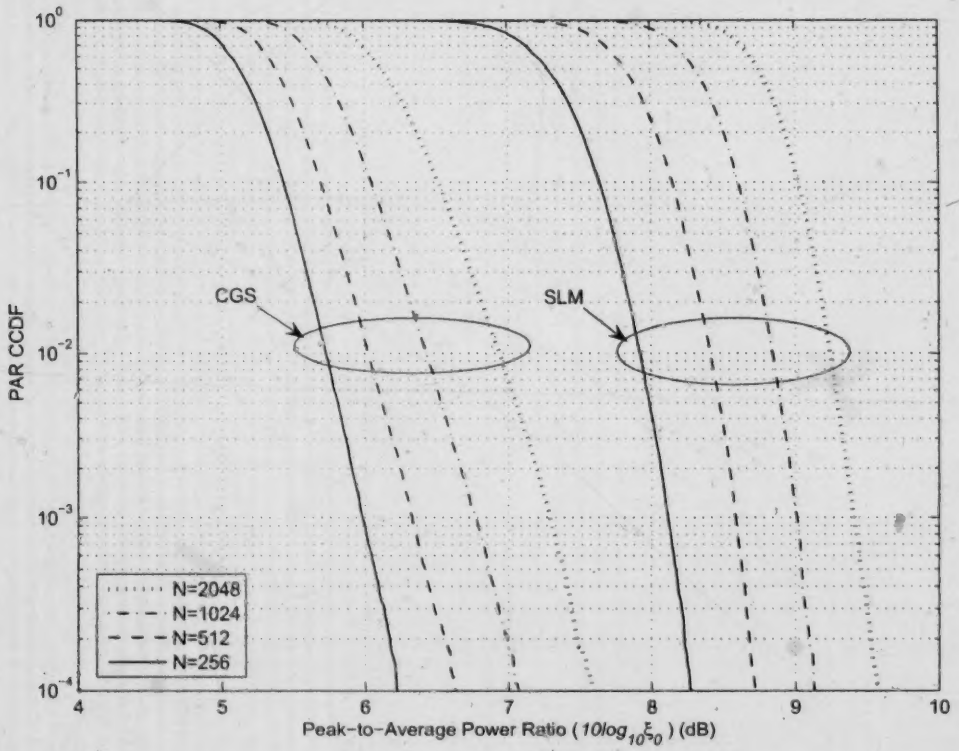


Figure 5.19: PAR reduction comparison of CGS and SLM for different N, where $K = 16$ for SLM, $L = 8$ and $A = 4$ dB for CGS, and 16QAM symbol input is used.

5.6 Conclusions

In this chapter, we proposed an adaptive mapping scheme for the sign-selection PAR-reduction technique to eliminate side information. We also proposed several algorithms (based on using a stochastic search or the clipping noise as a guide) to solve the associated discrete optimization problem of the sign-selection technique. The complexity analysis and simulation confirmed the complexity advantages of the proposed algorithms compared to the selective mapping and the derandomization algorithms.

Chapter 6

Coding Technique

We generalize the Rudin-Shapiro sequences [24–26] and study their availability for combining the multiple signal representation techniques to further reduce the PAR.

6.1 Rudin-Shapiro Sequence for MPSK Symbols

The ordinary Rudin-Shapiro sequence is used only for BPSK. However, this sequence can be easily extended to encode n MPSK symbols:

$$p_{i+1}(z) = p_i(z) + z^{2^i} e^{j\phi_i} q_i(z), \tag{6.1}$$

$$q_{i+1}(z) = p_i(z) - z^{2^i} e^{j\phi_i} q_i(z), \tag{6.2}$$

where $p_0(z) = q_0(z) = 1$, and ϕ_i are MPSK symbols. The PAR of $p_n(z)$ or $q_n(z)$ is no larger than 2.

$p_n(z)$ and $q_n(z)$ can be written in matrix form in accordance with

$$\begin{aligned} p_n(z) &= \mathbf{z} e^{j\mathbf{p}_n}, \\ q_n(z) &= \mathbf{z} e^{j\mathbf{q}_n}, \end{aligned} \tag{6.3}$$

where $\mathbf{z} = \begin{bmatrix} 1 & z & \cdots & z^{N-1} \end{bmatrix}$, $e^{j\mathbf{p}_n} = \begin{bmatrix} e^{jp_0} & e^{jp_1} & \cdots & e^{jp_{N-1}} \end{bmatrix}^T$, $N = 2^n$ is the sequence length, and $\mathbf{p}_n$, $\mathbf{q}_n$ are the corresponding MPSK phase vectors. By defining the input symbol vector $\mathbf{u} = \begin{bmatrix} \phi_0 & \cdots & \phi_{n-1} \end{bmatrix}$, the encoding from $\mathbf{u}$ to $\mathbf{p}_n$ and $\mathbf{q}_n$ can be written as

$$\mathbf{p}_n = \mathbf{u}\mathbf{G} + \mathbf{r}_n^{(p)}, \tag{6.4}$$

$$\mathbf{q}_n = \mathbf{u}\mathbf{G} + \mathbf{r}_n^{(q)}, \tag{6.5}$$

where $\mathbf{G}$ is the $n \times N$ generating matrix, and $\mathbf{r}_n^{(p)}$ and $\mathbf{r}_n^{(q)}$ are the corresponding phase compensation vectors with

$$\mathbf{r}_0^{(p)} = \mathbf{r}_0^{(q)} = \left[0\right], \tag{6.6}$$

$$\mathbf{r}_n^{(p)} = \left[\mathbf{r}_{n-1}^{(p)} \quad \mathbf{r}_{n-1}^{(q)}\right], \tag{6.7}$$

$$\mathbf{r}_n^{(q)} = \left[\mathbf{r}_{n-1}^{(p)} \quad \overline{\mathbf{r}_{n-1}^{(q)}}\right]. \tag{6.8}$$

Here, $\overline{\mathbf{a}}$ represents the element-wise π's complementary (i.e., for each element of $\mathbf{a}$, $\overline{0} = \pi$ and $\overline{\pi} = 0$). The generation matrix $\mathbf{G}$ can be constructed such that its i-th row $\mathbf{G}_{i*}$ has the form

$$\mathbf{G}_{i*} = \underbrace{\left[\hat{\mathbf{g}}_i \; \hat{\mathbf{g}}_i \; \cdots\right]}_{2^{n-i+1} \text{ items}}, \quad i = 1, ..., n, \tag{6.9}$$

$$\hat{\mathbf{g}}_i = \left[\mathbf{0}_{1\times 2^{i-1}} \quad \mathbf{1}_{1\times 2^{i-1}}\right], \tag{6.10}$$

where $\mathbf{0}_{1\times 2^{i-1}}$ and $\mathbf{1}_{1\times 2^{i-1}}$ represent the length 2^{i-1} all 0 and all 1 row vectors, respectively. Like the coding rate of the original Rudin-Shapiro sequence, the coding rate is

$$R = \frac{n}{N} = \frac{n}{2^n}. \tag{6.11}$$

6.2 Further Generalization of Rudin-Shapiro Sequence

A further generalization of the Rudin-Shapiro Sequence can be obtained by extending $p_0(z)$ and $q_0(z)$ to length-m polynomials with each containing $m-1$ PSK symbols in accordance with

$$p_0(z) = 1 + ze^{j\phi_0} + z^2e^{j\phi_1} + \cdots + z^{m-1}e^{j\phi_{m-2}}, \tag{6.12}$$

$$q_0(z) = 1 + ze^{j\phi_{m-1}} + z^2e^{j\phi_m} + \cdots + z^{m-1}e^{j\phi_{2m-3}}, \tag{6.13}$$

where $m \geq 1$, and $e^{j\phi_k} \triangleq 0$, when $k < 0$ (i.e. $p_0(z) = q_0(z) = 1$ when $m = 1$). Correspondingly, the recursive generation functions are

$$p_{i+1}(z) = p_i(z) + z^{2^i m}e^{j\phi_{i+2m-2}}q_i(z), \tag{6.14}$$

$$q_{i+1}(z) = p_i(z) - z^{2^i m}e^{j\phi_{i+2m-2}}q_i(z). \tag{6.15}$$

We then find

$$\mathrm{PAR}(p_n(z)) = \frac{1}{N}\max_z |p_n(z)|^2 \le 2m. \tag{6.16}$$

The coding rate of this generalized Rudin-Shapiro sequence is

$$R = \frac{n+2m-2}{2^n m}. \tag{6.17}$$

The generation matrix $\mathbf{G}$ can be constructed recursively in accordance with

$$\mathbf{G} = \begin{cases} \begin{bmatrix} \mathbf{G}_{n-1} & \mathbf{G}_{n-1} \\ \mathbf{0}_{1\times(2^{n-1}m)} & \mathbf{1}_{1\times(2^{n-1}m)} \end{bmatrix}, & N > 1, \\ \begin{bmatrix} \mathbf{G}_0^{(p)} & \mathbf{G}_0^{(q)} \\ \mathbf{0}_{1\times m} & \mathbf{1}_{1\times m} \end{bmatrix}, & N = 1. \end{cases} \tag{6.18}$$

where

$$\mathbf{G}_0^{(p)} = \begin{bmatrix} \mathbf{0}_{(m-1)\times 1} & \mathbf{I}_{(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \mathbf{0}_{(m-1)\times(m-1)} \end{bmatrix}, \tag{6.19}$$

$$\mathbf{G}_0^{(q)} = \begin{bmatrix} \mathbf{0}_{(m-1)\times 1} & \mathbf{0}_{(m-1)\times(m-1)} \\ \mathbf{0}_{(m-1)\times 1} & \mathbf{I}_{(m-1)} \end{bmatrix}, \tag{6.20}$$

with $\mathbf{I}_{(m-1)}$ representing $(m-1)$-by-$(m-1)$ identity matrix.

The PAR of the generalized Rudin-Shapiro sequence is determined by only its initial sequence. Suppose the PAR of $p_0(z)$ and $q_0(z)$ is upbounded by PAR_0; then, $\mathrm{PAR}_{p_n} \le 2\mathrm{PAR}_0$. Therefore, by reducing the PAR of p_0 and q_0, the PAR of p_n can also be reduced.

6.3 Simulation Results

Fig. 6.1 shows the PAR CCDF of the generalized Rudin-Shapiro sequences, where the length of the coded sequences is $N = 128$. Different lengths of the initial sequences are simulated. The PAR is upbounded by $2m$. A larger m gives a larger coding rate, but leads to a larger PAR. When m is large ($m = 32$), the PAR CCDF is close to that of the uncoded OFDM sequences. On the other hand, if the optimal signs of the initial sequences are used, the PAR of the generalized Rudin-Shapiro sequence can be significantly reduced. For example, when $m = 8$, and the signs of initial sequences are not optimized, the PAR is about 10.5 dB at PAR CCDF of 10^{-4}. However, when optimal signs are used for the initial sequences, the PAR is reduced to only 6.8 dB.

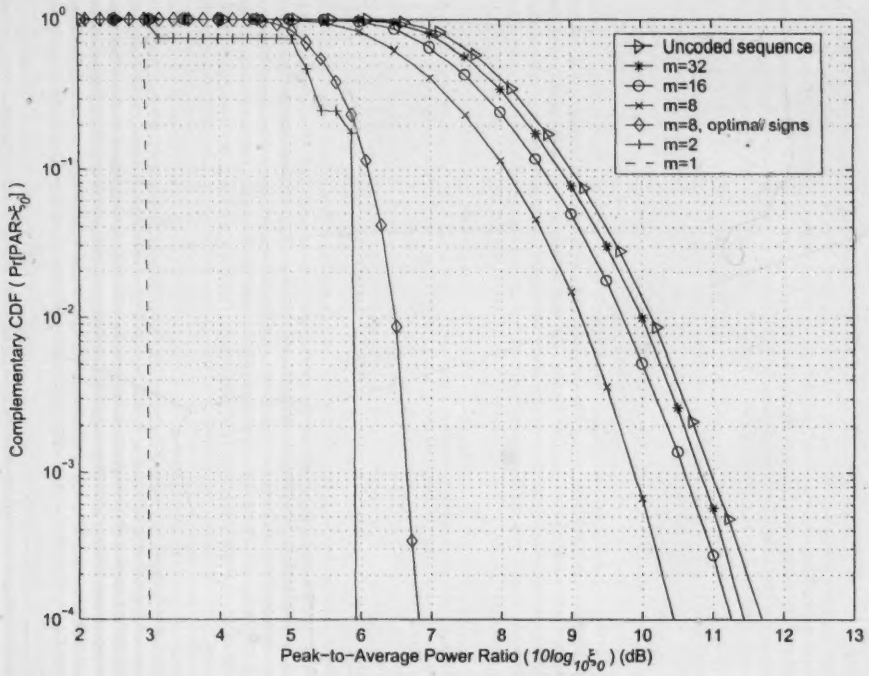


Figure 6.1: PAR CCDF of the generalized Rudin-Shapiro sequence for different m.

6.4 Conclusions

In this chapter, we generalized the Rudin-Shapiro sequences. Constructed from an initial PSK sequence, the generalized sequence increases the coding rate at the cost of an increased PAR. By optimizing the signs of the initial sequence, the PAR of the generalized sequence can be further reduced.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

In this thesis, we analyzed the PAR of OFDM systems and the amplitude clipping of OFDM signals with large peaks. Several PAR-reduction algorithms were proposed in Chapters 4 to 6.

Chapter 4 analyzed the clipping noise by approximating it as a series of parabolic pulses. Both the case in which the clipping noise consists of a single pulse and the case of multiple pulses have been discussed. Our analysis explained the peak regrowth and the constant clipping noise power spectrum over the whole OFDM band. We also established the roughly proportional relationship between the clipping noise at the end of several clipping and filtering iterations, and that generated in the first iteration. The constant of proportionality was estimated by using the level-crossing theory [15, 16].

In Chapter 4, a constant-scaling algorithm and an adaptive-scaling algorithm were proposed to reduce the PAR under the tone-reservation constraints. These algorithms scale the filtered first-iteration clipping noise by a constant or adaptively-calculated factor to compensate for peaks above the threshold. Analysis and simulation results showed that our proposed algorithms achieved a larger PAR reduction and lower complexity than the active-set algorithm. We also proposed a fast method to calculate the PAR and to find the clipping noise. This method can also be used in other PAR-reduction techniques such as SLM and PTS.

Chapter 5 focused on reducing the PAR by using the sign-selection technique. We first proposed an adaptive mapping scheme to eliminate the need for side information

at the receiver. Then, several algorithms were proposed to solve the discrete optimization problem associated with the sign-selection technique. The modified PTS algorithm first applied an initial sign sequences to the OFDM block, and then partitioned the OFDM block into subblocks and optimized the sign of each subblock to minimize the PAR. The recursive partial sequence algorithm also partitioned the OFDM block into subblocks. A suboptimal sign sequence was found by first minimizing the PAR of each subblock by optimizing the signs of the data symbols in this subblock, and then iteratively optimizing the sign of each subblock. The CESS algorithm used the CE method to solve the discrete optimization problem. The CGS algorithm used the clipping noise to guide the sign optimization. These algorithms offer a flexible tradeoff between PAR reduction and execution time, and obtain a larger PAR reduction than that obtained by existing techniques.

In Chapter 6, we generalized the Rudin-Shapiro sequences [24–26]. Constructed from an initial PSK sequence, the generalized sequence increased the coding rate at the cost of an increased PAR. By optimizing the signs of the initial sequence, the PAR of the generalized sequence can be further reduced.

7.2 Future Work

The search for fast algorithms to achieve large PAR reduction for OFDM systems still faces many challenges. Future research could be carried out on the following topics:

- The PAR distribution of the peak reduced OFDM signal for the low clipping threshold.

  In this thesis, we have derived the proportionality between the total clipping noise and the clipping noise generated in the first iteration. The PAR distribution of the peak reduced OFDM signal can then be derived for iterative clipping and filtering. However, when the clipping threshold is relatively low, the proportionality is biased by the interactions of the neighboring clipping pulses. Therefore, calculating the PAR distribution of the peak reduced OFDM signal is complicated. Future research on this topic would facilitate the evaluation and comparison of different clipping-based algorithms, and would help to develop new PAR-reduction algorithms to find the optimal solution.

- The optimal clipping threshold for clipping-based PAR-reduction algorithms.

  Hitherto, the optimal clipping threshold for clipping-based PAR-reduction algorithms have been determined by simulation. Our research in this thesis showed that the optimal clipping threshold is closely related to the strength of the time-domain sidelobes of the reserved tones. In other words, if the number of reserved tones is small and/or the reserved tones are consecutive (i.e., large sidelobes), the optimal clipping threshold should be large. If a large number of reserved tones are randomly selected, the clipping threshold should be low. A theoretical study of the optimal clipping threshold may need to identify the PAR distribution of the peak reduced OFDM signal.

- Power increase in tone-reservation and ACE.

  A side effect of tone-reservation and ACE is that the average power is increased after PAR reduction. Although such a power increase is small, it should still be minimized while reducing the PAR. We observed in Section 4.4 that the active-set algorithm gives rise to the smallest average power increase among the algorithms we compared. However, the PAR reduction of the active-set algorithm is inferior to that of the adaptive-scaling algorithm. Future research may combine the active-set and adaptive-scaling algorithms to minimize both the PAR and the average power increase.

- Low complexity algorithm for evaluating the PAR without IFFT.

  All sign-selection algorithms must calculate the time-domain samples and evaluate the PAR of a large number of sign-sequence candidates. In this thesis, we proposed a fast algorithm to calculate the PAR without computing the magnitudes of all the time-domain samples. However, calculating the time-domain samples of a sign sequence still requires an IFFT. This requirement is costly because of the need to calculate a large number of candidates. A low complexity algorithm for evaluating the PAR without IFFT would allow us to use more sign-sequence candidates to find a larger PAR reduction. Such an algorithm would also facilitate the use of clipping-based algorithms because they also require FFT/IFFT to compute the clipping noise in the time and frequency

domains.

- Fast CESS algorithms.

  The CESS algorithm proposed in this thesis still requires a large number of sign-sequence candidates to obtain a large PAR reduction, because a relatively large number of candidates are required in each iteration to ensure a small variance in estimating the PAR distribution. Future research may find a better probability-estimation algorithm that estimates the PAR distribution with acceptable variance by using a small number of candidates. Moreover, a deterministic version of CESS may outperforms the stochastic one proposed in this thesis.

- The relationship between the clipping noise and the optimal signs.

  Section 5.5 showed that the suboptimal signs for the sign-selection technique may be found by using the clipping noise. However, the relationship between the clipping noise and the optimal signs still requires further study. In this thesis, the suboptimal signs were found by rounding the magnitude of the clipping noise in the frequency domain to 0 or 2, and by minimizing the upper bound of the rounding error. A more accurate estimation of the rounding error may lead to faster convergence. Also, the rounding error is not directly related to the PAR. Thus, a more appropriate criterion needs to be derived to obtain a larger PAR reduction.

- Low PAR codes with a high coding rate and PAR reduction for MIMO OFDM systems.

- Peak reduction criteria using a more appropriate measure than the PAR.

  The purpose of peak reduction is to minimize the inband distortion and out-of-band radiation caused by the nonlinearity of HPA. A small PAR does not always imply a small inband distortion and out-of-band radiation. Peak-reduction criteria using other measures have also been studied based on simulations [238–240]. [239] observed that, when nonlinear amplification is allowed to some extent, the distribution of the envelope, rather than that of the PAR, is a more rele-

vant measure. A theoretical analysis and comparison of different peak-reduction criteria would help to develop more efficient peak-reduction algorithms.

Bibliography

[1] R. van Nee and R. Prasad, *OFDM for Wireless Multimedia Communications.* Boston: Artech House Publishers, March 2000.

[2] J. Oh and J. M. Cioffi, "Sub-band rate and power control for wireless OFDM systems," *60th IEEE Veh. Technol. Conf. VTC2004-Fall*, vol. 3, pp. 2011–2014, Sept. 26–29, 2004.

[3] X. Ma, G. Giannakis, and S. Barbarossa, "Non-data-aided frequency-offset and channel estimation in OFDM: and related block transmissions," in *Proc. IEEE Int. Conf. Commun. (ICC) 2001*, vol. 6, Helsinki, June 11–14, 2001, pp. 1866–1870.

[4] P. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, no. 10, pp. 2908–2914, Oct. 1994.

[5] M. Russell and G. L. Stüber, "Interchannel interference analysis of OFDM in a mobile environment," in *45th IEEE Veh. Technol. Conf. (VTC 95)*, vol. 2, Chicago, IL, July 25–28, 1995, pp. 820–824.

[6] K. Sathananthan and C. Tellambura, "Coding to reduce both PAR and PICR of an OFDM signal," *IEEE Commun. Lett.*, vol. 6, no. 8, pp. 316–318, Aug. 2002.

[7] ——, "Reducing intercarrier interference in OFDM systems by partial transmit sequence and selected mapping," in *6th Int. Symp. on DSP for Commun. Systems (DSPCS'2002)*, Manly-Sydney, Australia, Jan. 28–31, 2002, pp. 234–238.

[8] ——, "Partial transmit sequence and selected mapping schemes to reduce ICI in OFDM systems," *IEEE Commun. Lett.*, vol. 6, no. 8, pp. 313–315, Aug. 2002.

[9] ——, "Novel adaptive modulation scheme to reduce both PAR and ICI of an OFDM signal," in *6th Int. Symp. on DSP for Commun. Systems (DSPCS'2002)*, Manly-Sydney, Australia, Jan. 28–31, 2002, pp. 229–233.

[10] L. Wang and C. Tellambura, "A novel ICI cancellation technique for OFDM systems using adaptive mapping signal constellation," in *IEEE Pacific Rim Conf. on Commun., Comput., Signal processing (PACRIM) 2003*, vol. 1, Victoria, BC, Canada, Aug. 28-30, 2003, pp. 442–445.

[11] Y. Zhao and S.-G. Häggman, "Intercarrier interference self-cancellation scheme for OFDM mobile communication systems," *IEEE Trans. Commun.*, vol. 49, no. 7, pp. 1185–1191, July 2001.

[12] X. Li and L. J. Cimini, "Effects of clipping and filtering on the performance of OFDM," *IEEE Commun. Lett.*, vol. 2, no. 5, pp. 131–133, May 1998.

[13] S. H. Müller and J. B. Huber, "A novel peak power reduction scheme for OFDM," in *8th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, vol. 3, Helsinki, Sept. 1-4, 1997, pp. 1090–1094.

[14] J. Davis and J. Jedwab, "Peak-to-mean power control in OFDM, Golay complementary sequences, and Reed-Muller codes," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, pp. 2397–2417, Nov. 1999.

[15] N. M. Blachman, "Gaussian noise – part I: The shape of large excursions," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1396–1400, Nov. 1988.

[16] M. Pätzold, *Mobile Fading Channels*. John Wiley & Sons, Ltd., 2002.

[17] A. Bahai, M. Singh, A. Goldsmith, and B. Saltzberg, "A new approach for evaluating clipping distortion in multicarrier systems," *IEEE J. Select. Areas Commun.*, vol. 20, no. 5, pp. 1037–1046, June 2002.

[18] J. Tellado, "Peak to average power reduction for multicarrier modulation," Ph.D. dissertation, Stanford University, Sept. 1999.

[19] L. Wang and C. Tellambura, "A simplified clipping and filtering technique for PAR reduction in OFDM systems," *IEEE Signal Processing Lett.*, vol. 12, no. 6, pp. 453–456, June 2005.

[20] B. Krongold and D. Jones, "An active-set approach for OFDM PAR reduction via tone reservation," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 495–509, Feb. 2004.

[21] R. Rubinstein, *Simulation and the Monte Carlo Methods.* John Wiley and Sons, 1981.

[22] M. Sharif and B. Hassibi, "Existence of codes with constant PMEPR and related design," *IEEE Trans. Signal Processing*, vol. 52, pp. 2836–2846, Oct. 2004.

[23] ——, "Towards reducing the gap between PMEPR of multicarrier and single carrier signals," in *6th IEEE Workshop on Signal Processing advances in Wireless Commun.*, June 5–8, 2005, pp. 380–384.

[24] S. Boyd, "Multitone signals with low crest factor," *IEEE Trans. Circuits Syst.*, vol. 33, no. 10, pp. 1018–1022, Oct. 1986.

[25] W. Rudin, "Some theorems on Fourier coefficients," *Proc. Amer. Math. Soc.*, vol. 10, no. 6, pp. 855–859, Dec. 1959.

[26] H. S. Shapiro, "Extremal problems for polynomials and power series," Ph.D. dissertation, M.I.T., 1951.

[27] G. L. Stüber, *Principles of Mobile Communications*, 2nd ed. Norwell, MA, USA: Kluwer Academic Publishers, 2001.

[28] A. Jayalath, "OFDM for wireless broadband communications (peak power reduction, spectrum and coding)," Ph.D. dissertation, School of Computer Science and Software Engineering, Monash University, May 2002.

[29] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill Companies, Inc., 2001.

[30] C. Tellambura, "Use of m-sequence for OFDM peak-to-average power ratio reduction," *IEE Elect. Lett.*, vol. 33, no. 15, pp. 1300–1301, July 1997.

[31] ——, "Phase optimisation criterion for reducing peak-to-average power ratio in OFDM," *IEE Elect. Lett.*, vol. 34, no. 2, pp. 169–170, Jan. 1998.

[32] H. Rowe, "Memoryless nonlinearities with Gaussian inputs: Elementary results," *Bell Syst. Tech. J.*, vol. 61, pp. 1519–1525, Sept. 1982.

[33] A. Saleh, "Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers," *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1715–1720, Nov. 1981.

[34] ——, "Intermodulation analysis of FDMA satellite systems employing compensated and uncompensated TWT's," *IEEE Trans. Commun.*, vol. 30, pp. 1233–1242, May 1982.

[35] C. Tellambura, "Upper bound on peak factor of N-multiple carriers," *IEE Elect. Lett.*, vol. 33, no. 19, pp. 1608–1609, 1997.

[36] R. van Nee and A. de Wild, "Reducing the peak-to-average power ratio of OFDM," in *48th IEEE Veh. Technol. Conf. (VTC 98)*, vol. 3, Ottawa, Ont., May 18–21, 1998, pp. 2072–2076.

[37] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinearAWGN channels," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 430–441, Mar. 2000.

[38] P. Banelli, G. Baruffa, and S. Cacopardi, "Effects of HPA nonlinearity on frequency multiplexed OFDM signals," *IEEE Trans. Broadcast.*, vol. 47, no. 2, pp. 123–136, June 2001.

[39] D. Dardari, V. Tralli, and A. Vaccari, "A theoretical characterization of nonlinear distortion effects in OFDM systems," *IEEE Trans. Commun.*, vol. 48, no. 10, pp. 1755–1764, Oct. 2000.

[40] N. Dinur and D. Wulich, "Peak to average power ratio in amplitude clipped high order OFDM," in *Proc. IEEE Military Commun. Conf. (MILCOM) 1998*, vol. 2, Boston, MA, Oct. 18–21, 1998, pp. 684–687.

[41] ——, "Peak-to-average power ratio in high-order OFDM," *IEEE Trans. Commun.*, vol. 49, no. 6, pp. 1063–1072, June 2001.

[42] S. Litsyn and A. Yudin, "Discrete and continuous maxima in multicarrier communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 919–928, Mar. 2005.

[43] S. Litsyn and G. Wunder, "Generalized bounds on the crest-factor distribution of OFDM signals with applications to code design," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 992–1006, Mar. 2006.

[44] H. Ochiai and H. Imai, "Channel capacity of clipped OFDM systems," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT) 2000*, Sorrento, June 25–30, 2000.

[45] ——, "Performance of the deliberate clipping with adaptive symbol selection for strictly band-limited OFDM systems," *IEEE J. Select. Areas Commun.*, vol. 18, no. 11, pp. 2270–2277, Nov. 2000.

[46] ——, "On the distribution of the peak-to-average power ratio in OFDM signals," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 282–289, Feb. 2001.

[47] H. Ochiai, "Power efficiency comparison of OFDM and single-carrier signals," in *56th IEEE Veh. Technol. Conf. (VTC 2002-Fall)*, vol. 2, Sept. 24–28, 2002, pp. 899–903.

[48] H. Ochiai and H. Imai, "Performance analysis of deliberately clipped OFDM signals," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 89–101, Jan. 2002.

[49] H. Ochiai, "Performance analysis of peak power and band-limited OFDM system with linear scaling," *IEEE Trans. Wireless Commun.*, vol. 2, no. 5, pp. 1055–1065, Sept. 2003.

[50] K. Panta and J. Armstrong, "Effects of clipping on the error performance of OFDM in frequency selective fading channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 668–671, Mar. 2004.

[51] H.-G. Ryu, J. S. Park, and J.-S. Park, "Threshold IBO of HPA in the predistorted OFDM communication system," *IEEE Trans. Broadcast.*, vol. 50, no. 4, pp. 425–428, Dec. 2004.

[52] M. Sharif, M. Gharavi-Alkhansari, and B. Khalaj, "On the peak-to-average power of OFDM signals based on oversampling," *IEEE Trans. Commun.*, vol. 51, no. 1, pp. 72–78, Jan. 2003.

[53] D. Wulich, N. Dinur, and A. Glinowiecki, "Level clipped high-order OFDM," *IEEE Trans. Commun.*, vol. 48, no. 6, pp. 928–930, June 2000.

[54] G. Wunder and H. Boche, "Upper bounds on the statistical distribution of the crest-factor in OFDM transmission," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 488–494, Feb. 2003.

[55] ——, "Peak value estimation of bandlimited signals from their samples, noise enhancement, and a local characterization in the neighborhood of an extremum," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 771–780, Mar. 2003.

[56] H. Wang and B. Chen, "Asymptotic distributions and peak power analysis for uplink OFDMA signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2004*, vol. 4, May 17–21, 2004, pp. [illegible]5–8.

[57] S. Leung, S. Ju, and G. Bi, "Algorithm for repeated clipping and filtering in peak-to-average power reduction for OFDM," *IEE Elect. Lett.*, vol. 38, no. 25, pp. 1726–1727, Dec. 2002.

[58] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," *IEE Elect. Lett.*, vol. 38, no. 5, pp. 246–247, Feb. 2002.

[59] N. Ermolova, "New companding transform for reduction of peak-to-average power ratio," in *56th IEEE Veh. Technol. Conf. (VTC 2002-Fall)*, vol. 3, Sept. 24–28, 2002, pp. 1404–1407.

[60] H. Gong, W. Ye, S. Feng, and F. Ke, "A threshold companding scheme for reducing peak-to-average power ratio of OFDM signals," in *Int. Conf. on Wireless Commun., Networking and Mobile Comput.*, vol. 1, Sept. 23–26, 2005, pp. 573–576.

[61] I. Hosseini, M. J. Omidi, K. Kasiri, A. Sadri, and P. G. Gulak, "Papr reduction in OFDM systems using polynomial-based compressing and iterative expanding," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2006*, vol. 4, 2006.

[62] X. Huang, J. Lu, J. Chuang, and J. Zheng, "Companding transform for the reduction of peak-to-average power ratio of OFDM signals," in *53rd IEEE Veh. Technol. Conf. (VTC 2001 Spring)*, vol. 2, Rhodes, May 6–9, 2001, pp. 835–839.

[63] X. Huang, J. Lu, and J. Zheng, "Reduction in PAPR of OFDM system using a revised companding," in *Int. Conf. on Commun. Systems (ICCS) 2002*, vol. 1, Nov. 25–28, 2002, pp. 62–66.

[64] X. Huang, J. Lu, J. Zheng, J. Chuang, and J. Gu, "Reduction of peak-to-average power ratio of OFDM signals with companding transform," *IEE Elect. Lett.*, vol. 37, no. 8, pp. 506–507, Apr. 12, 2001.

[65] X. Huang, J. Lu, J. Zheng, K. B. Letaief, and J. Gu, "Companding transform for reduction in peak-to-average power ratio of OFDM signals," *IEEE Trans. Wireless Commun.*, vol. 3, no. 6, pp. 2030–2039, Nov. 2004.

[66] T. Jiang, Y. Yang, and Y.-H. Song, "Exponential companding technique for PAPR reduction in OFDM systems," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 244–248, June 2005.

[67] ——, "Companding technique for PAPR reduction in OFDM systems based on an exponential function," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2005*, vol. 5, Nov. 28–Dec. 2, 2005.

[68] T. Jiang, W. Yao, P. Guo, Y. Song, and D. Qu, "Two novel nonlinear companding schemes with iterative receiver to reduce PAPR in multi-carrier modulation systems," *IEEE Trans. Broadcast.*, vol. 52, no. 2, pp. 268–273, June 2006.

[69] T. Jiang and G. Zhu, "Nonlinear companding transform for reducing peak-to-average power ratio of OFDM signals," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 342–346, Sept. 2004.

[70] C.-H. Liu, "Offset companding technique for peak-to-average power ratio reduction of OFDM signals," in *13th IEEE Workshop on Local and Metropolitan Area Networks, (LANMAN)*, Apr. 25–28, 2004, pp. 229–234.

[71] A. Mattsson, G. Mendenhall, and T. Dittmer, "Comments on: "reduction of peak-to-average power ratio of OFDM system using a companding technique"," *IEEE Trans. Broadcast.*, vol. 45, no. 4, pp. 418–419, Dec. 1999.

[72] T. G. Pratt, N. Jones, L. Smee, and M. Torrey, "OFDM link performance with companding for PAPR reduction in the presence of non-linear amplification," *IEEE Trans. Broadcast.*, vol. 52, no. 2, pp. 261–267, June 2006.

[73] G. Ren, H. Zhang, and Y. Chang, "A complementary clipping transform technique for the reduction of peak-to-average power ratio of OFDM system," *IEEE Trans. Consumer Electron.*, vol. 49, no. 4, pp. 922–926, Nov. 2003.

[74] A. Vallavaraj, B. Stewart, D. Harrison, and F. McIntosh, "Reduction of peak to average power ration of OFDM signals using companding," in *Int. Conf. on Commun. Systems (ICCS) 2004*, Sept. 6–8, 2004, pp. 160–164.

[75] C.-L. Wang and S.-J. Ku, "A low-complexity companding transform for peak-to-average power ratio reduction in OFDM systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2006*, vol. 4, 2006.

[76] X. Wang, T. T. Tjhung, and C. S. Ng, "Reduction of peak-to-average power ratio of OFDM system using a companding technique," *IEEE Trans. Broadcast.*, vol. 45, no. 3, pp. 303–307, Sept. 1999.

[77] ——, "Reply to the comments on "reduction of peak-to-average power ratio of OFDM system using a companding technique"," *IEEE Trans. Broadcast.*, vol. 45, no. 4, pp. 420–422, Dec. 1999.

[78] X. Wang, T. T. Tjhung, C. S. Ng, and A. A. Kassim, "On the SER analysis of a-law companded OFDM system," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2000*, vol. 2, San Francisco, CA, Nov. 27–Dec. 1, 2000, pp. 756–760.

[79] X. Wang, T. T. Tjhung, and Y. Wu, "On the SER and spectral analyses of a-law companded multicarrier modulation," *IEEE Trans. Veh. Technol.*, vol. 52, no. 5, pp. 1408–1412, Sept. 2003.

[80] G. Yang, J. Sun, and Y. Li, "An adaptive companding transform for OFDM systems," in *6th IEEE Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Commun.*, vol. 2, May 31–June 2, 2004, pp. 445–448.

[81] Z. Zhang, Y. Wu, and J. Hou, "An improved scheme of reducing peak-to-average power ratio in OFDM systems," in *6th IEEE Circuits and Systems Symposium on Emerging Technologies: Frontiers of Mobile and Wireless Commun.*, vol. 2, May 31–June 2, 2004, pp. 469–471.

[82] A. D. S. Jayalath and C. Tellambura, "Peak-to-average power ratio of IEEE 802.11a PHY layer signals," in *6th Int. Symp. on DSP for Commun. Systems (DSPCS'2002)*, Manly-Sydney, Australia, Jan. 28–31, 2002, pp. 31–36.

[83] A. Aggarwal and T. Meng, "Minimizing the peak-to-average power ratio of OFDM signals via convex optimization," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2003*, vol. 4, Dec. 1–5, 2003, pp. 2385–2389.

[84] S.-K. Deng and M.-C. Lin, "OFDM PAPR reduction using clipping with distortion control," in *Proc. IEEE Int. Conf. Commun. (ICC) 2005*, vol. 4, May 16–20, 2005, pp. 2563–2567.

[85] A. Saul, "Peak reduction for OFDM by shaping the clipping noise," *2004. VTC2004-Fall. 2004 IEEE 60th Vehicular Technology Conference*, vol. 1, pp. 443–447, Sept. 26–29, 2004.

[86] H. Chen and A. Haimovich, "Iterative estimation and cancellation of clipping noise for OFDM signals," *IEEE Commun. Lett.*, vol. 7, no. 7, pp. 305–307, July 2003.

[87] ——, "An iterative method to restore the performance of clipped and filtered OFDM signals," in *Proc. IEEE Int. Conf. Commun. (ICC) 2003*, vol. 5, May 11–15, 2003, pp. 3438–3442.

[88] J. Tellado, L. M. C. Hoo, and J. M. Cioffi, "Maximum likelihood detection of nonlinearly distorted multicarrier symbols by iterative decoding," in *IEEE Global Telecommn. Conf. (GLOBECOM) 1999*, vol. 5, 1999, pp. 2493–2498.

[89] ——, "Maximum-likelihood detection of nonlinearly distorted multicarrier symbols by iterative decoding," *IEEE Trans. Commun.*, vol. 51, no. 2, pp. 218–228, Feb. 2003.

[90] A. Gatherer and M. Polley, "Controlling clipping probability in DMT transmission," in *31st Asilomar Conf. on Signals, Systems and Computers*, vol. 1, Pacific Grove, CA, Nov. 2–5, 1997, pp. 578–584.

[91] H. Saeedi, M. Sharif, and F. Marvasti, "Clipping noise cancellation in OFDM systems using oversampled signal reconstruction," *IEEE Commun. Lett.*, vol. 6, no. 2, pp. 73–75, Feb. 2002.

[92] D. Kim and G. L. Stüber, "Clipping noise mitigation for OFDM by decision-aided reconstruction," *IEEE Commun. Lett.*, vol. 3, no. 1, pp. 4–6, Jan. 1999.

[93] A. Gusmao and R. Dinis, "Iterative receiver techniques for cancellation of deliberate nonlinear distortion in OFDM-type transmission," in *Int. OFDM Workshop04*, Dresden, Sept. 2004.

[94] A. Gusmao, R. Dinis, and P. Torres, "Low-PMEPR OFDM transmission with an iterative receiver technique for cancellation of nonlinear distortion," in *62nd IEEE Veh. Technol. Conf. (VTC 2005-Fall)*, vol. 4, Sept. 25–28, 2005, pp. 2367–2371.

[95] L. Xia, Z. Ying, T. Youxi, and L. Shaoqian, "Iterative estimation and cancellation of clipping distortion for turbo receiver in MIMO-OFDM system," in *2005 Int. Conf. on Commun., Circuits and Systems*, vol. 1, May 27–30, 2005, pp. 185–188.

[96] D. Dardari, V. Tralli, and A. Vaccari, "A novel low complexity technique to reduce non-linear distortion effects in OFDM systems," in *9th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, vol. 2, Boston, MA, Sept. 8–11, 1998, pp. 795–800.

[97] R. Dinis and A. Gusmao, "A class of nonlinear signal-processing schemes for bandwidth-efficient OFDM transmission with low envelope fluctuation," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 2009–2018, Nov. 2004.

[98] W. Henkel and V. Zrno, "PAR reduction revisited: An extension to Tellado's method," in *6th Int. OFDM-Workshop*, Hamburg Germany, Sept. 18–19, 2001, pp. 31–1–31–6.

[99] W. Henkel, V. Azis, and S. Trautmann, "The analytic treatment of the error probability due to clipping - a solved problem?" in *Int. Symp. on Inform. Theory and its Applications (ISITA 2004)*, Parma, Italy, Oct. 10–13, 2004.

[100] E. Lawrey and C. Kikkert, "Peak to average power ratio reduction of OFDM signals using peak reduction carriers," in *5th Int. Symposium on Signal processing and its Applications (ISSPA)*, vol. 2, Brisbane, Qld., Australia, Aug. 22–25, 1999, pp. 737–740.

[101] S. Thompson, J. Proakis, and J. Zeidler, "The effectiveness of signal clipping for PAPR and total degradation reduction in OFDM systems," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2005*, vol. 5, Nov. 28–Dec. 2, 2005.

[102] C.-L. Wang, Y. Ouyang, and H.-C. Chen, "A low-complexity peak-to-average power ratio reduction technique for OFDM-based systems," in *60th IEEE Veh. Technol. Conf. (VTC 2004-Fall)*, vol. 6, Sept. 26–29, 2004, pp. 4380–4384.

[103] L. Wang and C. Tellambura, "An adaptive-scaling tone reservation algorithm for PAR reduction in OFDM systems," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2006*, San Francisco, CA, U.S.A., Nov. 27 – Dec. 1, 2006.

[104] ——, "Analysis of clipping noise and tone-reservation algorithms for peak reduction in OFDM systems," *IEEE Trans. Veh. Technol.*, (Accepted 2007).

[105] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY: Cambridge University Press, Mar. 2004.

[106] B. Krongold and D. Jones, "PAR reduction in OFDM via active constellation extension," *IEEE Trans. Broadcast.*, vol. 49, no. 3, pp. 258–268, Sept. 2003.

[107] Z. Yang, H. Fang, and C. Pan, "ACE with frame interleaving scheme to reduce peak-to-average power ratio in OFDM systems," *IEEE Trans. Broadcast.*, vol. 51, no. 4, pp. 571–575, Dec. 2005.

[108] S. H. Müller and J. B. Huber, "A comparison of peak power reduction schemes for OFDM," in *IEEE Global Telecommn. Conf. (GLOBECOM) 1997*, vol. 1, Nov. 3–8, 1997, pp. 1–5.

[109] P. Pushkarev, K.-W. Ryu, K.-Y. Yoo, and Y.-W. Park, "A study on the PAR reduction by hybrid algorithm based on the PTS and SLM techniques," in *57th IEEE Veh. Technol. Conf. (VTC 2003-Spring)*, vol. 2, Apr. 22–25, 2003, pp. 1263–1267.

[110] M. Rodrigues and I. Wassell, "IMD reduction with SLM and PTS to improve the error-probability performance of nonlinearly distorted OFDM signals," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 537–548, Mar. 2006.

[111] G. Zhou, R. Baxley, and N. Chen, "Selected mapping with monomial phase rotations for peak-to-average power ratio reduction in OFDM," in *Int. Conf. on Commun., Circuits and Systems (ICCCAS 2004)*, vol. 1, June 27–29, 2004, pp. 66–70.

[112] G. Zhou and L. Peng, "Optimality condition for selected mapping in OFDM," *IEEE Trans. Signal Processing*, vol. 54, no. 8, pp. 3159–3165, Aug. 2006.

[113] C. Tellambura and A. Jayalath, "PAR reduction of an OFDM signal using partial transmit sequences," in *54th IEEE Veh. Technol. Conf. (VTC 2001-Fall)*, vol. 1, Atlantic City, NJ USA, 2001, pp. 465–469.

[114] Y. Xin and I. J. Fair, "Low complexity PTS approaches for PAPR reduction of OFDM signals," in *Proc. IEEE Int. Conf. Commun. (ICC) 2005*, vol. 3, May 16-20, 2005, pp. 1991–1995.

[115] T. T. Nguyen and L. Lampe, "On partial transmit sequences to reduce PAR in OFDM systems," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2006*, San Francisco, CA, USA, Nov. 2006, pp. 1–6.

[116] F. Elguibaly, "A fast parallel multiplier-accumulator using the modified booth algorithm," *IEEE Trans. Circuits Syst. II*, vol. 47, no. 9, pp. 902–908, Sept. 2000.

[117] D.-W. Lim, S.-J. Heo, J.-S. No, and H. Chung, "On the phase sequence set of SLM OFDM scheme for a crest factor reduction," *IEEE Trans. Signal Processing*, vol. 54, no. 5, pp. 1931–1935, May 2006.

[118] R. Bauml, R. Fischer, and J. Huber, "Reducing the peak-to-average power ratio of multicarrier modulation by selected mapping," *IEE Elect. Lett.*, vol. 32, no. 22, pp. 2056–2057, Oct. 1996.

[119] Y. C. Cho, S. H. Han, and J. H. Lee, "Selected mapping technique with novel phase sequences for PAPR reduction of an OFDM signal," in *60th IEEE Veh. Technol. Conf. (VTC2004-Fall)*, vol. 7, Sept. 26–29, 2004, pp. 4781–4785.

[120] S. W. Kim, H. S. Byeon, J. K. Kim, and H.-G. Ryu, "An SLM-based real-time PAPR reduction method using dummy sequence insertion in the OFDM communication," in *Information, Communications and Signal Processing, 2005 Fifth International Conference on*, Dec. 06–09, 2005, pp. 258–262.

[121] A. Jayalath, C. Tellambura, and H. Wu, "Reduced complexity PTS and new phase sequences for SLM to reduce PAP of an OFDM signal," in *51st IEEE Veh. Technol. Conf. (VTC 2000-Spring)*, vol. 3, Tokyo, May 15–18, 2000, pp. 1914–1917.

[122] A. Jayalath and C. Tellambura, "SLM and PTS peak-power reduction of OFDM signals without side information," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2006–2013, Sept. 2005.

[123] S. H. Müller and J. B. Huber, "OFDM with reduced peak-to-average power ratio by optimum combination of partial transmit sequences," *IEE Elect. Lett.*, vol. 33, no. 5, pp. 368–369, Feb. 1997.

[124] L. Yang, R. Chen, Y. Siu, and K. Soo, "PAPR reduction of an OFDM signal by use of PTS with low computational complexity," *IEEE Trans. Broadcast.*, vol. 52, no. 1, pp. 83–86, Mar. 2006.

[125] M. Sharif and B. Hassibi, "On multicarrier signals where the PMEPR of a random codeword is asymptotically $\log n$," *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 895–903, May 2004.

[126] L. J. Cimini, Jr. and N. R. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences," *IEEE Commun. Lett.*, vol. 4, no. 3, pp. 86–88, Mar. 2000.

[127] B. M. Lee and R. de Figueiredo, "A low complexity tree algorithm for PTS-based PAPR reduction in wireless OFDM," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2006*, vol. 4, 2006, pp. IV–301–IV–304.

[128] S. H. Han and J. H. Lee, "PAPR reduction of OFDM signals using a reduced complexity PTS technique," *IEEE Signal Processing Lett.*, vol. 11, no. 11, pp. 887–890, Nov. 2004.

[129] G. Lu, P. Wu, and C. Carlemalm-Logothetis, "Peak-to-average power ratio reduction in OFDM based on transformation of partial transmit sequences," *IEE Elect. Lett.*, vol. 42, no. 2, pp. 105–106, Jan. 2006.

[130] W. S. Ho, A. Madhukumar, and F. Chin, "Peak-to-average power reduction using partial transmit sequences: a suboptimal approach based on dual layered phase sequencing," *IEEE Trans. Broadcast.*, vol. 49, no. 2, pp. 225–231, June 2003.

[131] A. Alavi, C. Tellambura, and I. Fair, "PAPR reduction of OFDM signals using partial transmit sequence: an optimal approach using sphere decoding," *IEEE Commun. Lett.*, vol. 9, no. 11, pp. 982–984, Nov. 2005.

[132] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, pp. 2806–2818, Aug. 2005.

[133] ——, "On the expected complexity of sphere decoding," in *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, vol. 2, Pacific Grove, CA, USA, 2001, pp. 1051–1055.

[134] W. H. Mow, "Universal lattice decoding: principle and recent advances," *Wireless Communications and Mobile Computing*, vol. 3, no. 5, pp. 553–569, Aug. 2003.

[135] H. Vikalo and B. Hassibi, "On the sphere-decoding algorithm II. Generalizations, second-order statistics, and applications to communications," *IEEE Trans. Signal Processing*, vol. 53, pp. 2819–2834, Aug. 2005.

[136] B. Wu, S. Cheng, and H. Wang, "Trellis factor search PTS for PAPR reduction in OFDM," in *16th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, vol. 4, Sept. 11–14, 2005, pp. 2514–2517.

[137] D.-W. Lim, J.-S. No, C.-W. Lim, and H. Chung, "A new SLM OFDM scheme with low complexity for PAPR reduction," *IEEE Signal Processing Lett.*, vol. 12, no. 2, pp. 93–96, Feb. 2005.

[138] D.-W. Lim, S.-J. Heo, J.-S. No, and H. Chung, "A new PTS OFDM scheme with low complexity for PAPR reduction," *IEEE Trans. Broadcast.*, vol. 52, no. 1, pp. 77–82, Mar. 2006.

[139] C.-L. Wang, M.-Y. Hsu, and Y. Ouyang, "A low-complexity peak-to-average power ratio reduction technique for OFDM systems," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2003*, vol. 4, Dec. 1–5, 2003, pp. 2375–2379.

[140] C.-L. Wang and Y. Ouyang, "A low-complexity selected mapping scheme for peak-to-average power ratio reduction in OFDM systems," in *60th IEEE Veh. Technol. Conf. (VTC 2004-Fall)*, vol. 1, Sept. 26–29, 2004, pp. 665–668.

[141] ——, "Low-complexity selected mapping schemes for peak-to-average power ratio reduction in OFDM systems," *IEEE Trans. Signal Processing*, vol. 53, no. 12, pp. 4652–4660, Dec. 2005.

[142] H. Qian, C. Xiao, N. Chen, and G. Zhou, "Dynamic selected mapping for OFDM," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2006*, vol. 4, Mar. 18–23, 2005.

[143] C. Tellambura, "Improved phase factor computation for the PAR reduction of an OFDM signal using PTS," *IEEE Commun. Lett.*, vol. 5, no. 4, pp. 135–137, Apr. 2001.

[144] H. Chen and G. Pottie, "An orthogonal projection-based approach for PAR reduction in OFDM," *IEEE Commun. Lett.*, vol. 6, no. 5, pp. 169–171, May 2002.

[145] A. D. S. Jayalath and C. Tellambura, "Side information in PAR reduced PTS-OFDM signals," in *14th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, vol. 1, Sept. 7–10, 2003, pp. 226–230.

[146] L. Cimini, Jr. and N. Sollenberger, "Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences with embedded side information," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2000*, vol. 2, San Francisco, CA, Nov. 27–Dec. 1, 2000, pp. 746–750.

[147] C.-C. Feng, Y.-T. Wu, and C.-Y. Chi, "Embedding and detection of side information for peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences," *58th IEEE Veh. Technol. Conf. (VTC 2003-Fall)*, vol. 2, pp. 1354–1358, Oct. 6–9, 2003.

[148] S. H. Han and J. H. Lee, "Modified selected mapping technique for PAPR reduction of coded OFDM signal," *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 335–341, Sept. 2004.

[149] L. Wang and C. Tellambura, "A novel PAR reduction technique for OFDM systems using adaptive mapping," in *15th Int. Conf. on Wireless Commun.*, Calgary, AB, Canada, July 7–9, 2003, pp. 270–277.

[150] S. H. Han, J. M. Cioffi, and J. H. Lee, "Tone injection with hexagonal constellation for peak-to-average power ratio reduction in OFDM," *IEEE Commun. Lett.*, vol. 10, no. 9, pp. 646–648, Sept. 2006.

[151] A. Pezeshk and B. H. Khalaj, "Extended hexagonal constellations as a means of multicarrier PAPR reduction," in *Proceedings of the First EurAsian Conference on Information and Communication Technology (EurAsia-ICT '02)*, vol. Volume 2510/2002. Shiraz, Iran: Springer-Verlag, Oct. 29-31, 2002, pp. 926–936.

[152] G. D. Forney, Jr, "Trellis shaping," *IEEE Trans. Inform. Theory*, vol. 38, no. 2, pp. 281–300, Mar. 1992.

[153] W. Henkel and B. Wagner, "Another application for trellis shaping: PAR reduction for DMT (OFDM)," *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1471–1476, 2000.

[154] A. Mobasher and A. K. Khandani, "Integer-based constellation-shaping method for PAPR reduction in OFDM systems," *IEEE Trans. Commun.*, vol. 54, no. 1, pp. 119–127, Jan. 2006.

[155] H. Ochiai, "A novel trellis-shaping design with both peak and average power reduction for OFDM systems," *IEEE Trans. Commun.*, vol. 52, no. 11, pp. 1916–1926, Nov. 2004.

[156] M. Vaezi, A. Mirzaee, and S. Safavi, "Power-efficient m-QAM signal constellations to reduce the PMEPR in OFDM systems," in *16th IEEE Int. Symp. on Personal, Indoor and Mobile Radio Commun. (PIMRC)*, vol. 2, Sept. 11–14, 2005, pp. 1135–1139.

[157] A. D. S. Jayalath and C. Tellambura, "Reducing the peak-to-average power ratio of orthogonal frequency division multiplexing signal through bit or symbol interleaving," *IEE Elect. Lett.*, vol. 36, pp. 1161–1163, June 22, 2000.

[158] ——, "The use of interleaving to reduce the peak-to-average power ratio of an OFDM signal," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2000*, vol. 1, San Francisco, CA, Nov. 27–Dec. 1, 2000, pp. 82–86.

[159] G. Lu, P. Wu, and C. Carlemalm-Logothetis, "Enhanced interleaved partitioning PTS for peak-to-average power ratio reduction in OFDM systems," *IEE Elect. Lett.*, vol. 42, no. 17, pp. 983–984, Aug. 2006.

[160] Z. bin Zhao, Y.-H. Kim, and J.-M. Kim, "PAPR reduction of an OFDM signal using sub-optimal PTS combined with guided scrambling," in *IEEE Int. Symposium on Commun. and Inform. Technol. (ISCIT 2005)*, vol. 1, Oct. 12–14, 2005, pp. 768–771.

[161] P. Van Eetvelt, G. Wade, and M. Tomlinson, "Peak to average power reduction for OFDM schemes by selective scrambling," *IEE Elect. Lett.*, vol. 32, no. 21, pp. 1963–1964, 1996.

[162] S. Sumathi, "Peak to average power ratio reduction of OFDM signal," in *2005 Annual IEEE INDICON*, Dec. 11–13, 2005, pp. 241–244.

[163] H. Breiling, S. Muller-Weinfurtner, and J. Huber, "SLM peak-power reduction without explicit side information," *IEEE Commun. Lett.*, vol. 5, no. 6, pp. 239–241, June 2001.

[164] Y. Xin and I. J. Fair, "Peak-to-average power ratio reduction of an OFDM signal using guided scrambling coding," in *IEEE Global Telecommn. Conf. (GLOBECOM) 2003*, vol. 4, Dec. 1-5, 2003, pp. 2390–2394.

[165] S. Ohno and G. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 12, pp. 2113–2123, Dec. 2002.

[166] N. Chen and G. Zhou, "Peak-to-average power ratio reduction in OFDM with blind selected pilot tone modulation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2210–2216, Aug. 2006.

[167] A. Jones, T. Wilkinson, and S. Barton, "Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission schemes," *IEE Elect. Lett.*, vol. 30, no. 25, pp. 2098–2099, Dec. 1994.

[168] H. Ochiai, "Analysis and reduction of peak-to-average power ratio in OFDM systems," Ph.D. dissertation, The Graduate School of Engineering, The University of Tokyo, Mar. 2001.

[169] K. Paterson, "Generalized Reed-Muller codes and power control in OFDM modulation," *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 104–120, Jan. 2000.

[170] Y. Xin and I. J. Fair, "Multiple-shift complementary sequences and their peak-to-average power ratio values," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT) 2004*, June 27–July 2, 2004, p. 121.

[171] C. Rößing and V. Tarokh, "A construction of OFDM 16-QAM sequences having low peak powers," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 2091–2094, July 2001.

[172] H. Lee and S. Golomb, "A new construction of 64-QAM Golay complementary sequences," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1663–1670, Apr. 2006.

[173] Z. Latinović and Y. Bar-Ness, "SFBC MIMO-OFDM peak-to-average power ratio reduction by polyphase interleaving and inversion," *IEEE Commun. Lett.*, vol. 10, no. 4, pp. 266–268, Apr. 2006.

[174] H. Lee, D. Liu, W. Zhu, and M. Fitz, "Peak power reduction using a unitary rotation in multiple transmit antennas," in *Proc. IEEE Int. Conf. Commun. (ICC) 2005*, vol. 4, May 16–20, 2005, pp. 2407–2411.

[175] Y.-L. Lee, Y.-H. You, W.-G. Jeon, J.-H. Paik, and H.-K. Song, "Peak-to-average power ratio in MIMO-OFDM systems using selective mapping," *IEEE Commun. Lett.*, vol. 7, no. 12, pp. 575–577, Dec. 2003.

[176] J.-H. Moon, Y.-H. You, W.-G. Jeon, K.-W. Kwon, and H.-K. Song, "Peak-to-average power control for multiple-antenna HIPERLAN/2 and IEEE802.11a systems," *IEEE Trans. Consumer Electron.*, vol. 49, no. 4, pp. 1078–1083, Nov. 2003.

[177] J. Shen, H. Fujii, T. Asai, and H. Yoshino, "Sequential search method with different thresholds for clipped power for OFDM signal," in *63rd IEEE Veh. Technol. Conf. (VTC 2006-Spring)*, vol. 4, 2006, pp. 2037–2041.

[178] M. Tan, Z. Latinović, and Y. Bar-Ness, "STBC MIMO-OFDM peak-to-average power ratio reduction by cross-antenna rotation and inversion," *IEEE Commun. Lett.*, vol. 9, no. 7, pp. 592–594, July 2005.

[179] S. O. Rice, "Distribution of the maxima of a random curve," *Amer. J. Math.*, vol. 61, pp. 409–416, 1939.

[180] ——, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 23, pp. 282–332, 1944.

[181] ——, "Mathematical analysis of random noise," *Bell Syst. Tech. J.*, vol. 24, pp. 46–156, 1945.

[182] ——, "Statistical properties of a sine wave plus random noise," *Bell Syst. Tech. J.*, vol. 27, pp. 109–157, 1948.

[183] ——, "Distribution of the duration of fades in radio transmission: Gaussian noise model," *Bell Syst. Tech. J.*, vol. 37, pp. 581–635, 1958.

[184] R. Price, "A useful theorem for nonlinear devices having Gaussian inputs," *IEEE Trans. Inform. Theory*, vol. 4, no. 2, pp. 69–72, June 1958.

[185] ——, "Comments on 'a useful theorem for nonlinear devices having Gaussian inputs'," *IEEE Trans. Inform. Theory*, vol. 10, no. 2, pp. 171–171, Apr. 1964.

[186] I. Blake and W. Lindsey, "Level-crossing problems for random processes," *IEEE Trans. Inform. Theory*, vol. 19, no. 3, pp. 295–315, May 1973.

[187] N. M. Blachman, "Gaussian noise – part II: Distribution of phase change of narrow-bandnoise plus sinusoid," *IEEE Trans. Inform. Theory*, vol. 34, no. 6, pp. 1401–1405, Nov. 1988.

[188] N. Blachman, "Gaussian noise: prediction based on its value and N derivatives," *IEE Proc. F Radar and Signal Processing*, vol. 140, no. 2, pp. 98–102, Apr. 1993.

[189] J. Mazo, "Asymptotic distortion spectrum of clipped, DC-biased, Gaussian noise [optical communication]," *IEEE Trans. Commun.*, vol. 40, no. 8, pp. 1339–1344, Aug. 1992.

[190] K. Sharpe, "Some properties of the crossings process generated by a stationary χ^2 process," *Advances in Applied Probability*, vol. 10, no. 2, pp. 373–391, June 1978.

[191] M. Aronowich and R. J. Adler, "Extrema and level crossings of χ^2 processes," *Advances in Applied Probability*, vol. 18, no. 4, pp. 901–920, Dec. 1986.

[192] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. Cambridge: Cambridge University Press, 1992.

[193] S. Bouguezel, M. O. Ahmad, and M. N. S. Swamy, "An efficient split-radix FFT algorithm," in *IEEE Int. Symp. on Circuits and System (ISCAS)*, vol. 4, May 25–28, 2003.

[194] ——, "A new radix-2/8 FFT algorithm for length-$q \times 2^m$ DFTs," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 9, pp. 1723–1732, Sept. 2004.

[195] ——, "Arithmetic complexity of the split-radix FFT algorithms," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2005*, vol. 5, Mar. 18–23, 2005, pp. V–137–V–140.

[196] S. C. Chan and P. M. Yiu, "An efficient multiplierless approximation of the fast fourier transform using sum-of-powers-of-two (SOPOT) coefficients," *IEEE Signal Processing Lett.*, vol. 9, no. 10, pp. 322–325, Oct. 2002.

[197] P. Duhamel, "Implementation of "split-radix" FFT algorithms for complex, real, and real-symmetric data," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 2, pp. 285–295, Apr. 1986.

[198] ——, "Algorithms meeting the lower bounds on the multiplicative complexity of length-2n DFTs and their connection with practical algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 9, pp. 1504–1511, Sept. 1990.

[199] H. Guo and C. S. Burrus, "Fast approximate Fourier transform via wavelets transform," in *Proc. SPIE Intl. Soc. Opt. Eng.*, M. A. Unser, A. Aldroubi, and A. F. Laine, Eds., vol. 2825, Oct. 23 1996, pp. 250–259.

[200] ——, "Wavelet transform based fast approximate Fourier transform," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Munich, Apr. 21-24, 1997, pp. 1973–1976.

[201] M. Heideman and C. Burrus, "On the number of multiplications necessary to compute a length-2^n DFT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 1, pp. 91–95, Feb. 1986.

[202] S. G. Johnson and M. Frigo, "A modified split-radix FFT with fewer arithmetic operations," *IEEE Trans. Signal Processing*, vol. 55, no. 1, pp. 111–119, Jan. 2007.

[203] S. Oraintara, Y.-J. Chen, and T. Nguyen, "Integer fast fourier transform (INTFFT)," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP) 2001*, vol. 6, Salt Lake City, UT, May 7–11, 2001, pp. 3485–3488.

[204] S. Oraintara, Y. J. Chen, and T. Q. Nguyen, "Integer fast fourier transform," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 607–618, Mar. 2002.

[205] M. A. Richards, "On hardware implementation of the split-radix FFT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 10, pp. 1575–1581, Oct. 1988.

[206] A. N. Skodras and A. G. Constantinides, "Efficient computation of the split-radix FFT," *IEE Proc. F Radar and Signal Processing*, vol. 139, pp. 56–60, Feb. 1992.

[207] H. Sorensen, M. Heideman, and C. Burrus, "On computing the split-radix FFT," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 1, pp. 152–156, Feb. 1986.

[208] H. Sorensen, D. Jones, M. Heideman, and C. Burrus, "Real-valued fast fourier transform algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 6, pp. 849–863, June 1987.

[209] D. Takahashi, "An extended split-radix FFT algorithm," *IEEE Signal Processing Lett.*, vol. 8, no. 5, pp. 145–147, May 2001.

[210] W.-C. Yeh and C.-W. Jen, "High-speed and low-power split-radix FFT," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 864–874, Mar. 2003.

[211] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover, 1964.

[212] L. Wang and C. Tellambura, "PAR reduction using adaptive mapping and recursive partial sequence method," in *3rd IASTED Int. Conf. on Wireless and Optical Commun.*, Banff, AB, Canada, July 14–16, 2003, pp. 654–658.

[213] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 13, 1983.

[214] A. Dekkers and E. Aarts, "Global optimization and simulated annealing," *Mathematical Programming*, vol. 50, pp. 367–393, Mar. 1991.

[215] H. E. Romeijn and R. L. Smith, "Simulated annealing for constrained global optimization," *Journal of Global Optimization*, vol. 5, no. 2, pp. 101–126, Sept. 1994.

[216] B. W. Wah and T. Wang, "Simulated annealing with asymptotic convergence for nonlinear constrained global optimization," *Principles and Practice of Constraint Programming -?CP'99*, pp. 461–475, 1999.

[217] F. Glover, "Tabu search – part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, Summer 1989.

[218] ——, "Tabu search – part II," *ORSA Journal on Computing*, vol. 2, no. 1, pp. 4–32, Winter 1990.

[219] B. L. Fox, "Integrating and accelerating tabu search, simulated annealing, and genetic algorithms," *Annals of Operations Research*, vol. 41, no. 2, pp. 47–67, June 1993.

[220] J.-K. Hao, R. Dorne, and P. Galinier, "Tabu search for frequency assignment in mobile radio networks," *Journal of Heuristics*, vol. 4, no. 1, pp. 47–62, June 1998.

[221] A. H. Wright, "Genetic algorithms for real parameter optimization," in *Foundations of genetic algorithms*, G. J. Rawlins, Ed. San Mateo, CA: Morgan Kaufmann, 1991, pp. 205–218.

[222] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, pp. 65–85, 1994.

[223] S. Khuri, T. Bäck, and J. Heitkötter, "An evolutionary approach to combinatorial optimization problems," in *Proceedings of the 1994 Computer Science Conference (CSC'94)*. Phoenix, Arizona: ACM Press, 1994, pp. 66–73.

[224] N. Chaiyaratana and A. Zalzala, "Recent developments in evolutionary and genetic algorithms: theory and applications," in *2nd Int. Conf. on Genetic Algorithms in Eng. Systems: Innovations and Applications, (GALESIA) 97*, Glasgow, Sept. 2–4, 1997, pp. 270–277.

[225] R. Rubinstein, "The cross-entropy method for combinatorial and continuous optimization," *Methodology and Computing in Applied Probability*, vol. 1, pp. 127–190, 1999.

[226] R. Y. Rubinstein, "Combinatorial optimization via cross-entropy," in *Encyclopedia of Operations Research and Management Sciences*, S. Gass and C. Harris, Eds. Kluwer, 2001, pp. 102–106.

[227] R. Rubinstein and D. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning*. Springer-Verlag, 2004.

[228] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of Operations Research*, vol. 134, no. 1, pp. 19–67, Feb. 2005.

[229] L. Margolin, "On the convergence of the cross-entropy method," *Annals of Operations Research*, vol. 134, no. 1, pp. 201–214, Feb. 2005.

[230] L. Wang and C. Tellambura, "Peak-to-average power ratio reduction of OFDM systems using cross entropy method," in *17th Int. Conf. on Wireless Commun.*, Calgary, AB, Canada, July 11–13, 2005, pp. 258–264.

[231] ——, "Peak-to-average power ratio reduction of OFDM systems using Cross Entropy method," submitted to *IEEE Trans. Signal Processing*, 2006 (3nd review).

[232] J. Stadler and S. Roy, "Adaptive importance sampling," *IEEE J. Select. Areas Commun.*, vol. 11, no. 3, pp. 309–316, Apr. 1993.

[233] S. Smith and G. Orsak, "A modified importance sampling scheme for the estimation of detection system performance," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1341–1346, 1995.

[234] P. Smith, M. Shafi, and H. Gao, "Quick simulation: a review of importance sampling techniques in communications systems," *IEEE J. Select. Areas Commun.*, vol. 15, no. 4, pp. 597–613, May 1997.

[235] M. Ferrari and S. Bellini, "Importance sampling simulation of turbo product codes," in *Proc. IEEE Int. Conf. Commun. (ICC) 2001*, vol. 9, Helsinki, June 11–14, 2001, pp. 2773–2777.

[236] S. Miller and R. O'Dea, "Peak power and bandwidth efficient linear modulation," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1639–1648, Dec. 1998.

[237] L. Wang and C. Tellambura, "A clipping guided sign-selection algorithm for PAR reduction OFDM systems," *IEEE Trans. Signal Processing*, (Accepted 2007).

[238] A. Behravan and T. Eriksson, "Some statistical properties of multicarrier signals and related measures," in *63rd IEEE Veh. Technol. Conf. (VTC 2006-Spring)*, vol. 4, Melbourne, Australia, May 7–10, 2006, pp. 1854–1858.

[239] H. Ochiai, "Amplifier efficiencies of OFDM systems with peak power and envelope variance reduction," in *IEEE Int. Symp. on Signal Processing and Inform. Technol. (ISSPIT)*, Aug. 2006, pp. 828–833.

[240] O. Takyu, T. Ohtsuki, and M. Nakagawa, "Criterion for Reducing Error Rate Degradation by Nonlinear Amplifier for Multicarrier Transmission," *IEICE Trans. Commun. (Japan)*, vol. E88-B, no. 7, pp. 3057–3061, 2005.